STATISTICAL METHODS AND APPLICATIONS TO ANIMAL BREEDING

by

ROBIN THOMPSON

B.Sc. (Newcastle)
M.Sc. (Newcastle)

Thesis presented for degree of Doctor of Science
University of Edinburgh

November 1987

EX BIBL. UNIV. EDINBURGH

CONTENTS

SUMMARY

This thesis comprises a collection of 39 research papers divided into three groups. The first group discusses the development of statistical methods, especially novel methods of variance component estimation, with general application. The second group examines the potential use of statistical methods in animal breeding studies, ranging from the construction of new experimental designs to the analysis of non-normal data. The third group reports on studies on animal breeding data in beef and dairy cattle.

Group I is entitled "Statistical methods, including variance component estimation, with general application". The major theme of this group is the estimation of variance components. Some previous work based on methods for balanced data, gave rise to methods that were neither unique nor efficient and other methods gave results that are inconsistent with the analysis of variance for balanced data. A method was introduced, now known as REML (Residual Maximum Likelihood) that unifies the area. The method was introduced for the analysis of incomplete block designs with unequal block size but was found to have important applications in the analysis of groups of similar trials, time-series and animal breeding. Papers investigating REML estimation for multivariate data, time-series and detecting outliers are included. The relationship of REML to other methods is elucidated, especially for balanced and partially balanced designs. Computational strategies are discussed.

The last two papers in the group illustrate a method of analysis of diallel crosses that involves using multiple copies of the data. This idea of using multiple copies was shown also to be useful in the analysis of rectangular lattice designs and in the interpretation of some recently introduced neighbour analyses of field trials.

The next group of papers, Group II, report on "Application of statistical methods to animal breeding studies". The work on variance components has some application in animal breeding and I have built on these links. Four papers consider efficient designs for estimation of genetic parameters, including designs for estimating heritability from data on two generations of data, for estimating maternal genetic variances, for estimating parent-offspring regression and for estimating multivariate genetic parameters. These designs can lead to substantial reductions in the variances of the estimates of the parameters, compared with classical designs, halving variances in some cases. Other papers have shown how to efficiently estimate heritability from unbalanced data, both from two generations of data and from more than two generations.

Often in animal breeding experiments animals used as parents are not selected at random, but selected on phenotypic measurements, perhaps of relatives. This can cause bias in some methods of estimation. On the other hand REML estimates can take account of the selection process. Selection experiments and the estimation of realised heritability are discussed.

REML estimation has found widespread acceptance by animal breeders, partly because some quantities arising in the

methods were terms that animal breeders use in evaluating animals. It was shown how to improve one method of evaluation and methods of evaluating sires were reviewed.

Some work is included on multivariate evaluation. It is shown how the complex multivariate calculations can be reduced to simpler univariate calculations using a canonical transformation, how results on selection indices can be used to interpret multivariate predictions. A simple interpretation of quadratic selection indices is given.

Other work considered some parallel problems with non-normal data. In particular for binary data, estimation of heritability, optimal designs for estimation of heritability and prediction of breeding values. It was shown how to estimate genotype frequencies using generalised linear model methods and suggested how to evaluate animals worth and estimate genetic parameters when the data fits a generalised linear model.

The last group, Group III, is entitled "Experimental studies". These include reports on a long term study of evaluation of breeds and cross-breeding in beef cattle in Zambia. The section also examines the genetic relationship between meat and milk production in British Friesian cattle. The validity of models used in dairy sire evaluation are investigated including the heterogeneity of heritability of milk yield at different levels of production and the use of a novel model for taking account of environmental variation within herds.

INDEX OF PAPERS

GROUP I STATISTICAL METHODS INCLUDING VARIANCE COMPONENT ESTIMATE WITH GENERAL APPLICATION

7. D.M. COOPER and R. THOMPSON. 1977. A note on the estimation of the parameters of the autoregressive-moving average process. Biometrika 64, 625-628.

8. R. THOMPSON. 1985. A note on restricted maximum likelihood estimation with an alternative outlier model. J.R. Statist. Soc. B 47, 53-55.

9. R. THOMPSON. 1975. A note on the W transformation. Technometrics 17, 511-512.

10. R. THOMPSON and K. MEYER. 1986. Estimation of variance components : what is missing in the EM algorithm? J. Statist. Comput. Simul. 24 215-230.

11. D.L. ROBINSON, R. THOMPSON and P.G.N. DIGBY. REML. 1982. A program for the analysis of non-orthogonal data by restricted maximum likelihood. COMPSTAT 1982, II. Eds. H. Cassinus, P. Ettinger and J.R. Mattieu. Physica-Verlag, Wien 231-232.

12. R. THOMPSON. 1983. Diallel crosses, partially balanced incomplete block designs with triangular association schemes and rectangular lattices. GENSTAT newsletter 10, 16-32.

13. R. THOMPSON. 1984. The use of multiple copies of data in forming and interpreting analysis of variance. Experimental design, Statistical Methods and Genetic Statistics. Ed. K. Hinkelmann. Marcel Dekker, New York, 155-174.

GROUP II APPLICATION OF STATISTICAL METHODS TO ANIMAL BREEDING STUDIES

14. R. THOMPSON. 1976. The estimation of maternal genetic variances. Biometrics 32 903-917.

15. R. THOMPSON. 1976. Design of experiments to estimate heritability when observations are available on parents and offspring. Biometrics 32 283-304.

16. W.G. HILL and R. THOMPSON. 1977. Design of experiments to estimate parent-offspring regression using selected parents. Anim. Prod. 24, 163-168.

17. N.D. CAMERON and R. THOMPSON. 1986. Design of multivariate selection experiments to estimate genetic parameters. Theor. Appl. Genet. 72, 466-476.

18. R. THOMPSON. 1977. The estimation of heritability with unbalanced data. I. Observations available on parents and offspring. Biometrics 33, 485-495.

19. R. THOMPSON. 1977. The estimation of heritability with unbalanced data. II. Data available on more than two generations. Biometrics 33, 495-504.

20. R. THOMPSON. 1977. The estimation of heritability with unbalanced data. III. Unpublished Appendices, 1-17.

21. R. THOMPSON. 1976. Estimation of quantitative genetic parameters. Proceedings of the International Conference on Quantitative Genetics. Ed. O. Kempthorne, E. Pollak and T. Bailey. Iowa State University press, Ames, Iowa, 639-657.

22. W.G. HILL and R. THOMPSON. 1978. Probabilities of non-positive definite between group or genetic covariance matrices. Biometrics 34, 429-439.

23. K. MEYER and R. THOMPSON. 1984. Bias in variance and covariance component estimators due to selection on a correlated trait. Z. Tierzucht. Zuchtungsbiol. 101, 33-50.

24. R. THOMPSON. 1976. Relationship between the cumulative different and best linear unbiased predictor methods of evaluating bulls. Anim. Prod. 23, 15-24.

25. R. THOMPSON. 1979. Sire Evaluation. Biometrics 35, 339-353.

26. R. THOMPSON. 1986. Estimation of realised heritability in a selected population using mixed model methods. Genet. Sel. Evol. 18, 475-484.

27. R. THOMPSON. 1972. The maximum likelihood approach to the estimate of liability. Anim. Hum. Genet. 36, 221-231.

28. R. THOMPSON, B.J. McGUIRK and A.R. GILMOUR. 1985. Estimating the heritability of all-or-none and categorical traits by offspring-parent regression. Z. Tierzucht. Zuchtungsbiol. 102, 342-354.

29. J.L. FOULLEY, D. GIANOLA and R. THOMPSON. 1983. Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty, birth weight and pelvic opening. Genet. Sel. Evol. 15, 401-424.

30. R. THOMPSON and R.J. BAKER. 1981. Composite link functions in generalised linear models. J.R. Statist. Soc. B. 30, 125-131.

31. R. THOMPSON. 1980. A note on the estimation of economic values for selection indices. Anim. Prod. 31, 115-117.

GROUP III EXPERIMENTAL STUDIES

32. W. THORPE, D.K.R. CRUICKSHANK and R. THOMPSON. 1980. Genetic and evironmental influences on beef cattle production in Zambia. Factors affecting weaner production from Angoni, Barotse and Boran dams. Anim. Prod. 30, 217-234.

33. W. THORPE, D.K.R. CRUICKSHANK and R. THOMPSON. 1980. Genetic and environmental influences on beef cattle production in Zambia. 2. Sire weights for age of purebred and reciprocally crossbred progeny. Anim. Prod. 30, 235-243.

34. W. THORPE, D.K.R. CRUICKSHANK and R. THOMPSON. 1980. Genetic and environmental influences on beef cattle production in Zambia. 3. Carcass characteristics of purebred and reciprocally crossbred progeny. Anim. Prod. 30, 245-252.

35. W. THORPE, D.C.K. CRUICKSHANK and R. THOMPSON. 1982. Genetic and environmental influences on beef cattle in Zambia. 4. Weaner production from purebred and reciprocally crossbred progeny. Anim. Prod. 33, 165-177.

36. W. THORPE, D.K.R. CRUICKSHANK and R. THOMPSON. 1979. The growth and carcass characteristics of crosses of Hereford and Friesian with Angoni, Barotse and Boran cattle in Zambia. J. Agric. Sci., Camb. 93, 423-430.

37. I.L. MASON, V.E. VIAL and R. THOMPSON. 1972. Genetic parameters of beef characteristics and the genetic relationship between meat and milk production in British Friesian cattle. Anim. Prod. 14, 135-148.

38. W.G. HILL, M.R. EDWARDS, M-K A. AHMED and R. THOMPSON. 1983. Heritability of milk yield and composition at different levels and variability of production. Anim. Prod. 36, 59-68.

39. V.P.S. CHAUHAN and R. THOMPSON. 1986. Dairy sire evaluation using a "rolling months" model. Z. Tierzucht. Zuchtungsbiol 103, 321-333.

REVIEW OF PAPERS

For convenience of exposition the papers have been classified into three broad groups. These papers represent some of my research when employed by the ARC Unit of Statistics (ARCUS) (1967-1983) and the AFRC Animal Breeding Research Organisation (ABRO) (1983-1986). The first group of papers roughly correspond to my first affiliation in that they consider the development of statistical methods that can be thought to have general application. The third group of papers correspond to my second affiliation in that they report on experimental studies involving animal breeding data. The second group concerning application of statistical methpds to animal breeding corresponds to an intersection of my two affiliations. The papers are presented, within groups, in a logical progression rather than in strict chronological order. The groupings are a little artificial in that the work has advanced in a non opportunistic way in all three groups together. With, for example, animal breeding data motivating research into statistical methods both for specific animal breeding application and for general application. The three groups are more interlocking strands knitted together to give a strong rope rather than three distinct threads.

Group I. Statistical methods, including variance component estimation, with general application

The first thirteen papers are concerned primarily with variance component estimation.

Paper 1 is a convenient introduction. At the time (1968) variance component estimation could be a rather ad hoc procedure with methods being suggested by analogy with methods that were appropriate for fixed effect models and balanced data. The paper by Cunningham and Henderson refered to in Paper 1 illustrates this. They had noted that efficient estimates of fixed effects in a mixed model arose from modification of least squares equations, set up when all the effects were assumed fixed. They suggested that terms arising in these modified equations could be used to estimate variance components. The modified block estimates were used in place of least square block estimates in an analysis of variance. Unfortunately, Cunningham and Henderson did not correctly derive the expectation of these quadratic forms in the analysis of variance. Paper 1 corrects these expectations. I am still surprised that Cunningham and Henderson, two eminent animal breeders, made these errors.

Paper 2 derived more efficient estimation procedures of variance components. The paper was motivated by incomplete block experiments and this influenced the title. This led to the paper being overlooked for some time. We should have chosen a more appropriate title to reflect its potential applications.

We used maximum likelihood (ML) methods to construct estimates. However there is a well known problem with ML methods in that they do not take account of the fact that the fixed effects are being estimated. By using the likelihood of error contrasts we were able to circumvent this difficulty and derive estimators of variance components that agreed with analysis of variance methods for balanced designs. As this method uses error contrasts or residuals, we now prefer to call it residual maximum likelihood (REML) but it is often called restricted maximum likelihood.

With hindsight, I think some points in the paper deserve more emphasis. The specification of the model in terms of the expectation of the observation vector and the variance matrix is a simpler and less confusing way to specify models rather than the more common way of expressing a linear model in terms of fixed and random effects (i.e. Paper 1). However this latter form has advantages for explaining the relevant computations.

The spectral decomposition was introduced as a mere technical device to allow the construction of a determinantal term related to the distribution of the error contrasts. The decomposition can be used to construct independent sums of squares whose expectations are functions of variance components and the latent roots (see Paper 4). This gives a simple interpretation to the estimation equations and can sometimes reduce the computational burden of finding estimates.

Paper 3 used REML estimation and looked in more detail at balanced designs. It extended the method to singular

variance-covariance matrices which arise, for example, when a randomisation model is used in experimental design. At the time (1973) Rao had recently introduced his Minimum Norm Quadratic Unbiased Estimates (MINQUE) based on minimising the norm (distance) between a variance matrix and its estimate. The criterion was not well understood by statisticians, but we found REML estimates could be related to MINQUE.

Paper 4 attempts to put several years of use of REML estimation of variance components into context. I identify three uses of mixed linear models, to estimate fixed effects, to estimate variance components and to predict random effects. The terms arising in REML variance component estimation have a role to play in the other uses with efficient estimators of fixed effects and predictors of random effects a natural by product. Further, some of the terms in REML estimation can be thought of as equating sums of squares of predicted values to their expectation.

I tried to forge links between unbalanced designs and balanced designs by thinking of the spectral decomposition as an extended analysis of variance. I introduced the idea of calculating approximate stratum variances to mimic expected mean squares tables in balanced analyses of variance. These ideas I have found invaluable in understanding the structure of non-orthogonal data sets.

I also pointed out why some Bayesian estimators have

unfortunate properties, in that they are likely to give estimates of zero for between group variance components in well designed experiments. However this type of estimator is still being advocated (for example Gianola, Foulley and Fernando, 1986).

Paper 5 considers the problem of estimation of multivariate components of variance. The study was motivated by the example in the later part of the paper when data is available on offspring-parent pairs and there is selection of parents. I am still surprised to the extent to which the multivariate analysis can be thought of as a natural extension of the univariate analysis, provided suitable matrix operations are defined. The symmetry of the variance matrixes certainly caused me many problems before I arrived at the simple equations presented in the paper. A natural question not answered in the paper is how much is gained by using a multivariate approach instead of a univariate approach (for variance components) or bivariate approach (for covariance components). When the model in section 2 is appropriate, then I would imagine that there was little to gain. However, when there is selection a multivariate approach is essential so that parameters in the unselected population can be estimated unbiasedly.

Paper 6 again considers multivariate data but in a different setting. It is assumed that different fixed effect models hold for different traits and there is no residual covariance between traits. This model was motivated by an

experiment were male and female animals were kept in different environments and there was interest in a covariance between a trait measured on the males and one measured on the females. The estimation procedure, although tedious, is straightforward and has no real surprises, except that it is much simpler than when a residual covariance is required. I underestimated the use to which this model would be put. It is often used in sheep breeding studies when male correlates of female reproduction are investigated. It can also be used to model genotype-environment interaction.

Paper 7 applies REML estimation in a different setting, to autoregressive-moving average model often used with time-series. The study arose out of simulations of a moving average process that gave rise to unsatisfactory maximum likelihood (ML) estimates. By using REML more reasonable estimates were found. The unsatisfactory nature of the ML estimates arose because ML did not take account of the estimation of the mean when estimating the variance parameters. This research has lead to the investigation of REML in other time-series settings, including the need for differencing in series and the identification of mixed spectra (Tunnicliffe Wilson, 1987).

Paper 8 investigates two alternative models for a single outlier. One model is that there is a slippage in the expected value of the observation. Another is that an outlier arises from an error term with increased variance. It had been found

that ML estimates for the position of the outlier could differ under the two models. It was shown that REML estimates for the position of the outlier agreed under the two models, a result in accord with my intuition.

Most of the preceding papers have used iterative schemes based on second differentials of the log-like likelihood to estimate variance parameters. Paper 9 investigates a different scheme. The EM algorithm suggested by Dempster, Laird and Rubin has become a commonly advocated algorithm for producing ML estimates. In the discussion of the original paper I pointed out that for variance component estimation the procedure can be slow to converge. On the other hand for balanced designs, estimates can be derived simply. We introduced alternative parameterisations based on the analysis of variance, that converge much faster. Examples are given using two and three components and multivariate designs. Essentially, by thinking about the structure of the data, and using a parameterisation so that the parameters are uncorrelated the computational effort can be reduced.

Papers 10 and 11 are both concerned with computational problems of variance estimation. A tranformation had been suggested that was useful in this estimation. In Paper 10 a simpler computational form, implicit in paper 2, was suggested that reduced that computation effort by up to three quarters. A computer program, REML, has been written by D. Robinson (1986) to allow the REML estimation of variance components, the estimation of fixed effects and prediction of random effects and is

widespread use. Paper 11 discusses the strategies and algorithms used in the program.

The last two papers in this group (Papers 12 and 13) have less to do with variance component estimation. They were motivated by attempts to use analysis of variance techniques in the computer program GENSTAT to analyse diallel cross data. Because each observation had two contributions from parents (i.e. male parent and female parent) the model could not be specified in GENSTAT. I found that if two copies of the data were used and associated the male parent with the first copy and the female parent with the second copy and introduced a blocking factor connecting the two copies of each observation I could construct a valid analysis of variance. The same idea can be used to produce analyses of some partially balanced incomplete block designs and rectangular lattice designs.

In Paper 13 the idea of using multiple copies is used to investigate neighbour analyses of field experiments. It is shown that some neighbour analyses can be thought of as taking multiple copies of the data and imposing different blocking schemes on each copy. This connection between neighbour analyses and incomplete block analyses, although a little contrived, I feel was worth making in order to show that the methods were not too different. I also hoped that optimum incomplete block designs could be adapted to give efficient designs for neighbour analysis but no one has taken up this suggestion.

Group II Application of statistical methods to animal breeding studies

The first four papers (14-17) of this group are concerned with designs for the estimation of genetic parameters.

Paper 14 used two suggested designs for the investigation of maternal genetic variances and shows how to efficiently estimate these genetic parameters. Previous attempts to estimate parameters efficiently had been quite tedious. For one of the designs, one suggested method would have required the calculation of 91 variances and covariances of covariances between relatives. In order to streamline the calculation, multivariate analyses of variance were calculated and it was shown how the parameter estimates could be easily found from these analyses of variance. Both designs used led to high sampling variances and high corelations between some of the estimates. It was shown how to fit a non-linear alternative model with fewer parameters.

Paper 15 considered the case when there is interest in estimating heritability and observations are available on parents and offspring. Information on heritability can be derived from parent-offspring regression and sib-covariances. It was known (a) how to combine this information when there was random mating of parents and (b) that the sampling variances of regression estimates could be reduced by selection and random mating of

parents. My first aim was to show how to incorporate sib covariance information when there was selection and assortative mating. I introduced a multivariate analysis of variance with three variates being sire, dam and offspring measurements and four sources of variation, namely within dams, between dams within sires, between sires and between animals without offspring. This allowed the likelihood to be easily calculated and maximised using the technique given in Paper 14.

I then investigated optimal designs. I found that the variances of heritability estimates could be roughly halved over designs were parents were not selected. The sib covariance information was more valuable at low heritabilities and led to increases in optimal family size over previously suggested designs.

I finally extended the method of estimation to multivariate data by the ingenious, if confusing, device of replacing scalar products in the univariate formulae by direct products. The average Biometrics reader deserved a more detailed explicit derivation.

Paper 16 was a natural follow up to Paper 15, concerned with the case when the relative numbers in parent and offspring generations were fixed and there was interest in selecting parents and estimating the offspring-parent regression. The optimal proportions of animals to use as parents were derived. The value of the estimated parameter influences the optimum but the designs are robust to poor estimates. Again, if

appropriate, the sib covariance information can be combined with the regression estimate but use of this sib information does not materially effect the optimum proportion.

Paper 17 investigates designs for estimating multivariate genetic parameters and takes up a suggestion made in Paper 15 (10 years earlier). The classifical experimental design for investigation of two traits X and Y is to carry out two selection experiments in parallel. In one experiment parents are selected for high and low X values and offspring measured for responses in X and Y and these responses compared with the selection differential for X. In the other experiment parents are selected for high and low Y values and these compared with the selection differential for Y. This ignores potential information on the Y parental values in the first and the X parental values in the second experiment. We showed how to use this extra information in the non-selected trait, how to use the individual values rather than averages, and investigated the use of two linear indices as selection criteria rather than X and Y.

It was found that using two phenotypically uncorrelated induces, say L_1 and L_2, to select parents is more efficient than selecting on X and Y when the aim is to minimise a generalised variance of the three genetic variances and covariances. The most efficient design would depend on knowing the genetic parameters but the L_1 and L_2 design is very simple, flexible and surprisingly efficient.

An alternative to carrying out two selection schemes is to use one scheme selecting 'extreme' animals on a quadratic

index $L_1^2+L_2^2$. This scheme can reduce variances of genetic parameters by a factor of approximately two compared to classical designs for uncorrelated traits, and with bigger gains for correlated traits. The optimal proportions of animals to use as parents again depend on the genetic parameters, but as in Paper 16 the designs are robust to the choice of parameters. Extensions to deal with selection and assortative mating of male and female parents and with more than two generations of animals are under development.

The next five papers (18-22) are mainly concerned with the estimation of heritability.

Papers 19, 20 and 21 are companion papers. Papers 19 and 20 originally had appendices which were refereed and initially accepted for publication, but because of pressure on journal space were later removed. For completeness these appendices are included as Paper 21.

Paper 19 indicates how to calculate REML estimates for a genetic model including additive and dominance variances when information is available on parents and offspring. A sequential approach is followed absorbing offspring and family effects in turn. This is similiar to the use of Reduced Animal Models (RAM) for breeding value prediction (Blair and Pollak, 1984). The algorithm mentioned in the discussion and appendix has been quite successful and is the basis of the computer program REML (discussed in Paper 11).

Paper 20 considers the same problem when data is available on more than two generations. A formula for the additive variance matrix (A) between individuals is derived using arguments about the passage of genes from one generation to the next and a surprisingly simple formula for the inverse of A is obtained. Henderson had independently derived the same formula by a more algebraic argument. Henderson's aim was to give better predictors of breeding values using information between relatives. There has been little validation to see if this model is appropriate. The formulation of Paper 19 allows easy extension to test, for example, if half-sib covariances are equal to dam-daughter covariances.

There were no comparisons with other commonly used methods. Firstly, because the relative efficiencies of the methods depends on the balance and orthogonality of the designs and I am unsure how to characterise this to give general advice. Secondly, I felt that the coherence of the whole approach giving efficient estimation of genetic parameters and fixed effects and predictors of breeding values was worth the effort especially in the cases considered in Paper 19 and the one sex case in Paper 20. I admit the two sex case is more computationally demanding.

Paper 21 is the first invited paper I presented at a conference. Then, I thought a conference paper should (i) review recent work, (ii) present new, perhaps unpublishable ideas, (iii) give ideas for future work. This I tried to do, although I would not be so ambitious nowadays.

In particular I reviewed the work in Papers 15 and 19. I then showed that the multivariate estimation procedure for q traits developed in Paper 5 could be simplified. The original procedure involved an analysis involving all q traits together. I pointed out that if the traits were transformed then the analysis simplified to q univariate analyses. This transformation was called a canonical transformation because it produced traits that were uncorrelated, both phenotypically and genetically. The application was described more fully in Meyer (1985). The canonical transformation is also computationally useful in predicting multivariate breeding values and, as I anticipated, in examining the effects of errors in parameters on the efficiency of selection indices (Hayes and Hill, 1980).

Paper 22 is again concerned with multivariate estimation. When genetic, G, and residual variance component matrices are estimated, then heritabilities of linear combinations can be found. One would like the estimates of heritabilities of all linear combinations to be positive or zero. Firstly, because heritabilities are by definition positive or zero. Secondly, if a selection index is constructed using a negative heritability estimate one would select the worst rather than the best animals. To ensure that all heritabilities are not less than zero requires the estimated G matrix to be non-negative definite. We calculated the probability that the estimate of G was non-negative definite for a range of

heritabilities and number of traits by a mixture of simulation and analysis. Again a canonical transformation was useful in expressing the probability in terms of the smallest number of parameters as possible. The probability increased quite dramatically as the number of traits increased. This emphasised the need to inspect genetic matrices to ensure they are positive definite.

The next five papers (23-28) discussed the effect of selection on estimation of parameters and prediction of breeding values.

In several of the previous papers (notably 5 and 15) the use of ML estimators has been advocated because it can take account of selection, giving estimates of variance parameters in the unselected population. Paper 23 considers a more complicated case which can arise with dairy cattle, when a decision for a cow to be kept for a second lactation depends on its first lactation record and other factors such as type or temperament might play an important role. We investigated the case when selection is on one trait correlated with yield. We identified cases when the analysis of yield alone will give unbiased estimates of variance parameters and derive general formula for bias in analysis of variance (AOV) and ML estimators. Simulation results are in good agreement with these results showing that for the set of genetic parameters considered, ML estimators are less biassed than AOV estimators. The bias involved depends on the correlations of the culling variate with

the yield traits. A dummy binary survival trait was introduced to see if one could estimate all these correlations between the culling variate and milk yield. Unfortunately, but not surprisingly, not all correlations could be estimated.

Paper 24 was concerned with sire evaluation. In Britain in 1952 a contempory comparison (CC) method was introduced to evaluate dairy bulls. The difference between daughters of a bull and contempories of the bulls daughters were used to predict a breeding value for a bull. No account is taken of the genetic merit of contempories. It was recognised that this could lead to biased evaluations, but the extra computing effort was not thought worthwhile given the amount of effective selection that was being carried out. By 1970 the situation had changed in that there was more effective selection being carried out and there were better computing facilities available.

Two methods of predicting breeding values, Henderson's Best Linear Unbiased Predictor (BLUP) method and a more intuitive cumulative difference (CD) method were compared in Paper 25. Both methods were improvements to the CC method in that they attempted to take account of genetic metit of contempories. It was shown how to improve the CD method. The BLUP method is now the standard way of predicting breeding values, but at the time it was not understood by all animal

breeders. This comparison with the CC method and the presentation in terms of algebra rather than in terms of matrices, included at the editor's insistence, certainly helped some animal breeders to understand the method.

Paper 25 was an invited paper on sire evaluatiuon in a special issue of Biometrics on "Perspectives in Biometry". We were encouraged to put recent work in perspective and raise unsolved problems. The paper continues the discussion on sire evaluation with the connection between BLUP, CC and selection indices explained.

I then go on to discuss the effect of selection on prediction. Henderson had introduced a conditional model to take account of selection being on a trait correlated with the trait in the model. I give an ingenious argument, using imaginary effects, to derive Henderson's results. Henderson then considered the case when the conditioning variable is a linear function of the observations, L'y, and showed that when this function is location invariant that BLUP under a conditional and a unconditional model are the same. Whilst a interesting mathematical result, I am not sure about its relevance for two reasons. Firstly, it is unlikely that experimenters or farmers select animals on the basis of a location invariant L'y. Secondly, in a likelihood approach both L'y and y, given L'y are used in the inference. In Henderson's approach only y given L'y is used to estimate fixed effects. Despite prompting, the questions I raise have not been answered in a satisfactory way. Perhaps this conflict cannot be resolved and is just one facet of

the differences between a repeated sampling paradigm and a likelihood framework (Cox and Hinkley, 1974).

I then go on to discuss the worth of using genetic relationships and show how breeding values can be decomposed into independent components which lead to a more sensible way of incorporating and interpreting group effects.

BLUP can be used to calculate genetic and environmental trends. I point out that these genetic trends need careful interpretation and in some circumstances are not based on direct evaluations of genetic merit, but can be merely a multiple of selection differentials. Paper 26 follows up this latter point. It has been suggested that BLUP could be used to estimate realised heritabilities in selection experiments without controls. First genetic worth is predicted with BLUP using an assumed value of heritability. Then the regression of predicted genetic worth on cumulative selection differential is used as an estimate of realised heritability. For two simple designs I showed that the regression is more a function of the assumed heritability rather than the value of heritability in the population.

Returning to Paper 25 the final section is concerned with extensions to linear models for normally distributed data. Nedder and Wedderburn had introduced the idea of generalised linear models (GLM), which unified a large area of statistics. The generalisation came in two ways. Firstly, they considered data to be generated by distributions in the exponential family, which includes Normal, Poisson, Gamma and Binominal

distributions. Secondly, they allowed the mean value of an observation to be 'linked' to a linear function of explanatory variables by a transformation. The models can be fitted by a weighted least squares algorithm. Nelder and Wedderburn's development was in terms of fixed effect models. I suggested a modification to deal with random effect models. Several different problems (including Papers 28, 29) have been found to fit into this general framework. Perhaps it was naive to expect this general framework to be quickly understood or to expect that there was no need for the publication of the many special cases! The suggested modification does involve approximations. In some settings this is satisfactory (for example Hoeschel, 1986). However for binary data with small family sizes it can lead to unsatisfactory estimates of variance components (Gilmour, Anderson and Rae, 1985). Further work is needed in this area.

The next group (27-31) of five papers continue this theme and are concerned with estimatation for non-normally distributed data and prediction of non-linear functions.

Papers 27 and 28 discuss binary data. We follow Falconer and assume that a binary trait might follow the usual laws of quantitative genetics if examined at the level of liability rather than expression.

Paper 27 shows how maximum likelihood estimates of heritability of liability can be found when a binary trait is measured on sibships. It also shows that, when costs are

proportional to observations made, that gains in efficiency can be made by choosing the sibships from which further observations are taken on the basis of observations on probands. The results parallel the results for quantitative traits in that there can be reductions of up to a half in the variance of the heritability of liability if sibships are selected appropriately.

Paper 28 again looks at estimation of heritability for binary traits, using parent-offapring regression and allowing for the fact that incidence might differ according to when the measurement was taken. The estimation was conveniently described in terms of a GLM. A clear exposition is given of the differences between using the binary and liability scales.

Paper 29 discusses the case of prediction of genetic merit when data is available on both quantitative and binary traits. I think of this paper as a natural extension of GLM to multivariate traits with random effects, and a similar algorithm is used. Curiously, the method has application to fixed effect models but we do not mention this. The extension comes from sequentially thinking of the quantitative data and then the qualitative data given the quantitative data. Perhaps the method is not quite as natural as it appears. My co-authors had used Bayesian methods to derive the results. I became involved because their arguments about incorporation of information from quantitative traits was incorrect. I corrected this using a more intuitive approximate likelihood approach. Then we jointly

derived the Bayesian argument in the paper. A method has a better chance of being useful if it can be derived in more than one way.

As the previous papers have shown, generalised linear models have proved a powerful, flexible tool for modelling data when there is a one-to-one correspondence between observation and linear predictor. There are, however, situations when one would like to link sums, or differences, of functions of linear predictors to observations and Paper 30 gives examples. In this paper we introduce the idea of a composite link function allowing the linking of several linear predictors to one observation. The computing algorithm introduced by Nelder and Wedderburn can be simply modified to deal with this extension. Of course, there are several ways of deriving estimates in these situations and our approach is only one of many ways of generalising linear models. For estimating gene frequencies, our procedure would involve constructing complete contingency tables of genotypes and specifying how the cells are combined to give observations. This is a simple and instructive way of specifying the appropriate models.

In some of the work, especially Paper 17, I have found geometrical arguments useful and Paper 31 is another amusing instructive example. I was concerned with prediction of non-linear profit function and plotting contours of equal profit showed that what had been assumed to be a maximum was a saddle-point.

Group III Experimental studies

The third group of papers are concerned with experimental studies and involve the analysis of cattle data. Papers 32-36 are concerned with two long term studies comparing purebred and crossbred beef cattle in Zambia. These were very thorough including comparison of weaner production, growth rates and carcass characteristics for four purebreds and 12 crossbreds. Even now, ten years later, there are only two other studies carried out in Africa that are as comprehensive as these.

Paper 32 examined reproductive performance and calf and dam weights for two indigenous breeds, Angoni and Barotse, and an introduced breed, Boran, under ranching conditions in Zambia. In particular the Angoni dams were found to have a higher calving percentage but lower calf weaning weight than the Barotae and Boran dams. Year had a major influence on calving and weaning percentage.

Papers 33-35 consider the performance of purebred progeny from these three breeds and also the six possible reciprocal crossbred progeny from these three breeds. Purebred cattle of the introduced Africander breed were also included. Papers 33 and 34 consider liveweights up to 3 1/2 years and carcass characteristics. There were interactions of genotype with year for liveweight but overall the progeny of the

introduced breeds, Africander and Boran, were heavier and had heavier carcasses than the progeny of the indigenous breeds, Angoni and Barotse. For carcass traits not related to size, the two zebu breeds, Angoni and Boran, were similar as were the two sanga breeds, Africander and Barotse. The zebu breeds had more fat cover than the sanga breeds. Heterosis was only shown in the Barotse/Boran cross (a sanga/zebu cross) increasing to over 5% for weight at the later ages and carcass weight, and of the order of 3% for linear carcase measurements. The Angoni breed had smaller maternal effects than the Barotse and Boran breeds.

The scope of inter-breed selection and cross-breeding for improvement of carcass characteristics in Zambia was clearly demonstrated.

Paper 35 considers weaner production from these same 10 genotypes used as dams. The introduced breeds, Africander and Barotse, had poorer calving rates than the indigenous breeds, Angoni and Boran. Heterosis estimates were about 9% for calving rate and weaning rate for crosses with the Boran and much smaller (1% and 4%) for Angoni/Barotse crosses. The fertility of dams dry at mating was much larger than that of those lactating and calved late in the season. There were large differences between breeds for calf and dam weights but heterosis estimates (of the dam) for these traits were small and insignificant.

Paper 36 considers weight and carcass traits for the

three pure breeds, Angoni, Barotse and Boran with crosses with the exotic Friesian and Hereford breeds. Rankings of the purebreeds were similar to those found in Papers 33 and 34. Crossbred progeny had heavier liveweight and carcase weights than the purebreeds (about 13% for the Boran comparison). There were few important carcass differences between Friesian and Hereford crosses. The use of exotic crosses is clearly advantageous for growth and carcase production in Zambia.

Paper 37 was concerned with the genetic relationships between beef characters and milk production. I was introduced to the data, long after the last animal was slaughtered, to help because some traits were not measured on all animals and because it was not clear what the effect of using several different slaughter criteria would have on the relationships. These circumstances added to the difficulty of the analysis and partly contributed to the size of the standard errors. This analysis certainly motivated me to try to design better animal breeding experiments.

Paper 38 presents estimates of heritability of milk yield and composition at different levels and variability of production. It was found that if herds were split into two groups, high and low, either on a) level of production or b) herd variance or c) herd coefficient of variation, heritability of yield (or log yield) was higher in the high group than the low group. The genetic correlations were close to 1 between high

and low levels for all traits. The implications for progeny testing of bulls and evaluating the genetic merit of bulls and cows was considered.

Paper 39 considered alternative ways of adjusting for environmental effects within herds for dairy cattle data. Most methods have estimated fixed effects of seasons found by grouping several months together. This paper examines the covariances between month effects within herds. It shows that in some British data the covariance between month effects decreases linearly as the lag between months increases. This linear covariance model can be simply fitted by introducing extra pseudo-month effects. Whilst a logical model, taking more account of 'neighbour' information, it was found to have little practical advantage over fitting random herd-year-season effects. The approach might be of more value for pig breeding data. Often exponential smoothing techniques are used to adjust data and breeding values are then calculated on the adjusted data ignoring the fact that adjustments have been made.

BIBLIOGRAPHY

H.T. BLAIR and E.J. POLLAK. 1984. Comparison of an animal model and an equivalent reduced animal model for computational efficiency using mixed model methodology. J. Animal Science 38, 1090-1096.

D.R. COX and D.V. HINKLEY. 1974. Theoretical Statistics. Chapman and Hall, London.

D. GIANOLA, J.L. FOULLEY and R.L. FERNANDO. 1986. Prediction of breeding values when variances are not known. Genet. Sel. Evol. 18, 485-498.

A.R. GILMOUR, R.D. ANDERSON and A.L. RAE. 1985. The analysis of binomial data in a generalised linear model. Biometrika 72, 593-599.

J.F. HAYES and W.G. Hill. 1980. A reparameterization of a genetic index to locate its sampling properties. Biometrics 36, 237-248.

J. HOSCHELE. 1986. Estimation of breeding values and variance components with quasi continuous data. Ph.D. Thesis Hohenheim.

K. MEYER. 1985. Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. Biometrics 41, 153-165.

D.L. ROBINSON. 1987. Estimation and use of variance components. The Statistician 36, 3-14.

G. TUNNICLIFFE WILSON. 1987. On the use of marginal likelihood in time series model estimation. Submitted to R.T. Statist. Soc. B.

STATEMENT OF AUTHORSHIP

I am sole author of eighteeen of the thirty-nine papers in this collection. For the papers reported under joint authorship, the work in most cases was closely collaborative and individual contributions cannot be reliably quantified. However I have tried to indicate my approximate contribution under the three headings: Initiation, including conception and planning, execution, completion.

	Percentage Contribution		
Paper	Initiation	Execution	Completion
2	40	50	40
3	45	50	40
6	0	40	60
7	80	30	40
10	90	40	60
11	80	20	40
16	40	40	40
17	70	40	60
22	40	50	40
23	60	30	35
28	80	60	60
29	0	30	10
30	90	80	80
32-36	30	*	40
37	0	*	30
38	30	25	25
39	70	30	35

* For papers 32-37 my contribution to the execution was to do all the statistical analyses.

The work has not been submitted for other degrees, with the following exceptions. Paper 17 reports work included in a Ph.D thesis by N.D. Cameron (University of Edinburgh, 1987). Paper 23 was based on part of a Ph.D thesis by K. Meyer (University of Edinburgh, 1983) and Paper 39 was based on part of a Ph.D thesis by V.P.S. Chauhan (University of Edinburgh, 1985).

ACKNOWLEDGEMENTS

My contribution to the work reported in this collection was conducted at the ARC Unit of Statistics, 1967-83, AFRC Animal Breeding Research Organisation, 1983-86, and the Statistical Laboratory, Iowa State University, Ames, Iowa, 1972-73. It is a pleasure to acknowledge the benefit from these colleagues and co-authors. As directors, David Finney and Roger Land both allowed me my head. At the beginning of my career Maurice Bichard, Bill Hill, Oscar Kempthorne, Desmond Patterson and Robin Plackett tried to point me in the right direction. In later years Kevin Atkins , Neil Cameron, Vijay Chauhan, Chris Glasbey, Karin Meyer and Bill Thorpe have risked my impatience and asked searching questions.

Robin Thompson 24/11/87

ROBIN THOMPSON

1

Iterative estimation of variance components for non-orthogonal data

by

R. Thompson

Reprinted from
BIOMETRICS
THE BIOMETRIC SOCIETY, Vol. 25, No. 4, December 1969

ITERATIVE ESTIMATION OF VARIANCE COMPONENTS FOR NON-ORTHOGONAL DATA

R. THOMPSON

Agricultural Research Council, Unit of Statistics, Edinburgh, Scotland

SUMMARY

Cunningham and Henderson [1968] presented an iterative procedure for the estimation of constants and variance components in data conforming to a mixed model. This method is shown to be faulty and corrections are suggested. For one example corrected estimates are shown to be very different from those of Cunningham and Henderson [1968]. The practical value of this method is discussed.

INTRODUCTION

Eisenhart [1947] introduced the term 'mixed model' to describe models useful in experiments where some effects, such as block or animal effects, can be thought of as random effects and other effects, for example treatments, are regarded as fixed. The estimation of variance components from these experiments is fully understood for various incomplete block arrangements with a high degree of symmetry (Nelder [1968]). For a general non-orthogonal design, however, difficulties arise and no simple known method is optimal under all conditions.

Lack of balance is very common in records on animals, especially those arising from studies of quantitative genetics. Experimenters are fortunate if families of animals are all of the same size; even if an experiment begins with reasonable symmetry, accidental losses may introduce non-orthogonality. Cunningham and Henderson [1968] have proposed a general method of estimation, using iterative calculations. An algebraic oversight has corrupted their formulae, giving an iterative process that will usually fail to converge to anything reasonable.

THE GENERAL ANALYSIS

As far as possible the notation will be that of Cunningham and Henderson [1968]. A general non-orthogonal design for treatments and blocks gives yields, $\mathbf{y}$, which may be represented by

$$\mathbf{y} = \mathbf{Xa} + \mathbf{Zb} + \mathbf{e}.$$

If there are N observations, $\mathbf{y}$ is an $N \times 1$ vector of yields. The matrices $\mathbf{X}$ and $\mathbf{Z}$ consist of known elements specifying the treatment and block structure; $\mathbf{X}$ is of size $N \times q$ and rank q', and $\mathbf{Z}$ is of size $N \times r$ and rank r'. The vector $\mathbf{a}$ is a $q \times 1$ matrix of constants representing the treatment effects. The vector $\mathbf{b}$ is

a $r \times 1$ matrix of random variables with means 0 and such that

$$E(\mathbf{bb}') = \mathbf{I}\sigma_b^2 .$$

(In this paper, **I** represents an identity matrix of appropriate size.)

The vector **e** is an $N \times 1$ matrix of random variables, representing the intra-block component of error, with

$$E(\mathbf{e}) = 0, \qquad E(\mathbf{ee}') = \mathbf{I}\sigma_e^2 .$$

var (**y**), the $N \times N$ variance-covariance matrix of **y**, is then given by

$$\text{var } (\mathbf{y}) = \mathbf{I}\sigma_e^2 + \mathbf{ZZ}'\sigma_b^2 .$$

Cunningham and Henderson's matrix W is a $N \times (q + r)$ composite design matrix of rank p' made up of the columns of **X** and **Z**. The forms of **X** and **Z** are such that the general mean $\mathfrak{y}$ can be estimated either as a linear function of the treatment effects **a**, or as a linear function of the blocks effects **b**; so the rank of W can be at most $q' + r' - 1$.

Table 1 shows the form of analysis of variance usual for such data. It should be noted that Table 1 holds generally even if treatments and blocks are completely confounded. In the most common design, that of a two-way classification with no confounding, q' is the number of treatments, r' the number of blocks, and $p' = q' + r' - 1$. In this particular case, the degrees of freedom for $R(\mathbf{a} \mid \mathfrak{y}, \mathbf{b})$ and $R(\mathbf{b} \mid \mathfrak{y}, \mathbf{a})$ in Table 1 simplify to $q' - 1$ and $r' - 1$.

Equating of the mean squares for '**b** eliminating $\mathfrak{y}$, **a**' and 'residual' to their expectations gives consistent estimates for σ_b^2 and σ_e^2. The $R(\mathfrak{y}, \mathbf{a})$, $R(\mathbf{b} \mid \mathfrak{y}, \mathbf{a})$, and $R(\mathfrak{y}, \mathbf{a}, \mathbf{b})$ of Table 1 correspond to the $R(\mathbf{a})$, $R(\mathbf{b})$, and $R(\mathbf{a}, \mathbf{b})$ of Cunningham and Henderson. In general, these estimates are not of maximum efficiency, since contrasts between treatments contain information about block effects.

If σ_b^2, σ_e^2 were known, minimization of

$$(\mathbf{y} - \mathbf{Xa})'(\mathbf{I}\sigma_e^2 + \mathbf{ZZ}'\sigma_b^2)^{-1}(\mathbf{y} - \mathbf{Xa})$$

TABLE 1

COMPUTING FORMS FOR MIXED MODEL ANALYSIS OF VARIANCE

SV	*DF*	*SS*	*EMS*
Total	N	$\mathbf{y}'\mathbf{y}$	
a elim. $\mathfrak{y}$, b	$p' - r'$	$\hat{\mathbf{a}}'[\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}] = R(\mathbf{a}\vert\mathfrak{y}, \mathbf{b})$	
$\mathfrak{y}$, b ign. a	r'	$\mathbf{y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = R(\mathfrak{y}, \mathbf{b})$	
$\mathfrak{y}$, a, b	p'	$R(\mathbf{a}\vert\mathfrak{y}, \mathbf{b}) + R(\mathfrak{y}, \mathbf{b}) = R(\mathfrak{y}, \mathbf{a}, \mathbf{b})$	
$\mathfrak{y}$, a ign. b	q'	$\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = R(\mathfrak{y}, \mathbf{a})$	
b elim. $\mathfrak{y}$, a	$p' - q'$	$R(\mathfrak{y}, \mathbf{a}, \mathbf{b}) - R(\mathfrak{y}, \mathbf{a}) = R(\mathbf{b}\vert\mathfrak{y}, \mathbf{a})$	$\sigma_e^2 + c\sigma_b^2$
Residual	$N - p'$	$\mathbf{y}'\mathbf{y} - R(\mathfrak{y}, \mathbf{a}, \mathbf{b})$	σ_e^2

where $\hat{\mathbf{a}} = [\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}[\mathbf{X}' - \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{y}$
and $c = [\text{tr } \mathbf{Z}'\mathbf{Z} - \text{tr } \mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}]/(p' - q')$.

would estimate $\mathbf{a}$ efficiently. The estimator $\tilde{\mathbf{a}}$ of $\mathbf{a}$ satisfies

$$\begin{aligned} \mathbf{X}'\mathbf{X}\tilde{\mathbf{a}} + \mathbf{X}'\mathbf{Z}\tilde{\mathbf{b}} &= \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{X}\tilde{\mathbf{a}} + (\mathbf{Z}'\mathbf{Z} + k\mathbf{I})\tilde{\mathbf{b}} &= \mathbf{Z}'\mathbf{y}, \end{aligned} \tag{1}$$

where $k = \sigma_e^2/\sigma_b^2$.

Cunningham and Henderson [1968] based their method on these equations, calculating sums of squares as though (1) were normal equations. They regarded

$$R^*(\mathbf{y}, \mathbf{a}, \mathbf{b}) = \tilde{\mathbf{a}}'\mathbf{X}'\mathbf{y} + \tilde{\mathbf{b}}'\mathbf{Z}'\mathbf{y}$$

as a sum of squares due to fitting $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$. Now

$$\begin{aligned} &R^*(\mathbf{y}, \mathbf{a}, \mathbf{b}) \\ &= [\tilde{\mathbf{a}}, \tilde{\mathbf{b}}]'\begin{bmatrix}\mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y}\end{bmatrix} = [\mathbf{y}'\mathbf{X}, \mathbf{y}'\mathbf{Z}]\begin{bmatrix}\mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + k\mathbf{I}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y}\end{bmatrix} \\ &= \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &\quad + \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Z}\{k\mathbf{I} + \mathbf{Z}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Z}\}^{-1}\mathbf{Z}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} \\ &= R^*(\mathbf{y}, \mathbf{a}) + R^*(\mathbf{b} \mid \mathbf{y}, \mathbf{a}) \end{aligned}$$

(or again $R(\mathbf{a}) + R(\mathbf{b})$ in Cunningham and Henderson's notation).

$R^*(\mathbf{y}, \mathbf{a}, \mathbf{b})$ can also be written in the form

$$R^*(\mathbf{y}, \mathbf{a}, \mathbf{b}) = R^*(\mathbf{a} \mid \mathbf{y}, \mathbf{b}) + R^*(\mathbf{y}, \mathbf{b}),$$

where

$$R^*(\mathbf{a} \mid \mathbf{y}, \mathbf{b}) = \tilde{\mathbf{a}}'\mathbf{X}'[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-1}\mathbf{Z}']\mathbf{y}$$

and

$$R^*(\mathbf{y}, \mathbf{b}) = \mathbf{y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-1}\mathbf{Z}'\mathbf{y}.$$

To simplify the formulae let

$$\mathbf{S} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

and

$$\mathbf{T} = \{k\mathbf{I} + \mathbf{Z}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Z}\}^{-1},$$

so

$$\mathbf{T} = (\mathbf{Z}'\mathbf{S}\mathbf{Z} + k\mathbf{I})^{-1}.$$

Then

$$R^*(\mathbf{y}, \mathbf{a}) = R(\mathbf{y}, \mathbf{a}) = \mathbf{y}'(\mathbf{I} - \mathbf{S})\mathbf{y}$$

and

$$R^*(\mathbf{b} \mid \mathbf{y}, \mathbf{a}) = \mathbf{y}'\mathbf{S}\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{S}\mathbf{y}.$$

Cunningham and Henderson interpret $R^*(\mathbf{b} \mid \mathbf{y}, \mathbf{a})$ as a sum of squares due to $\tilde{\mathbf{b}}$ and $\mathbf{y}'\mathbf{y} - R^*(\mathbf{y}, \mathbf{a}, \mathbf{b})$ as a residual error sum of squares.

It should be noted that $R^*(\mathbf{y}, \mathbf{a}, \mathbf{b})$ is not the usual sum of squares found by fitting $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$. In fact,

$$
\begin{aligned}
&(\mathbf{y} - \mathbf{X}\tilde{\mathbf{a}} - \mathbf{Z}\tilde{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\tilde{\mathbf{a}} - \mathbf{Z}\tilde{\mathbf{b}}) \\
&\quad = \mathbf{y}'\mathbf{y} - \begin{bmatrix}\tilde{\mathbf{a}}\\ \tilde{\mathbf{b}}\end{bmatrix}'\begin{bmatrix}\mathbf{X}'\mathbf{y}\\ \mathbf{Z}'\mathbf{y}\end{bmatrix} - [\mathbf{y}'\mathbf{X} \quad \mathbf{y}'\mathbf{Z}]\begin{bmatrix}\tilde{\mathbf{a}}\\ \tilde{\mathbf{b}}\end{bmatrix} + \begin{bmatrix}\tilde{\mathbf{a}}\\ \tilde{\mathbf{b}}\end{bmatrix}'\begin{bmatrix}\mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}\\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{a}}\\ \tilde{\mathbf{b}}\end{bmatrix} \\
&\quad = \mathbf{y}'\mathbf{y} - 2\begin{bmatrix}\tilde{\mathbf{a}}\\ \tilde{\mathbf{b}}\end{bmatrix}'\begin{bmatrix}\mathbf{X}'\mathbf{y}\\ \mathbf{Z}'\mathbf{y}\end{bmatrix} + \begin{bmatrix}\tilde{\mathbf{a}}\\ \tilde{\mathbf{b}}\end{bmatrix}'\begin{bmatrix}\mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}\\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z}\end{bmatrix}\begin{bmatrix}\mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}\\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + k\mathbf{I}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}'\mathbf{y}\\ \mathbf{Z}'\mathbf{y}\end{bmatrix} \\
&\quad = \mathbf{y}'\mathbf{y} - 2\begin{bmatrix}\tilde{\mathbf{a}}\\ \tilde{\mathbf{b}}\end{bmatrix}'\begin{bmatrix}\mathbf{X}'\mathbf{y}\\ \mathbf{Z}'\mathbf{y}\end{bmatrix} + \begin{bmatrix}\tilde{\mathbf{a}}\\ \tilde{\mathbf{b}}\end{bmatrix}'\begin{bmatrix}\mathbf{X}'\mathbf{y}\\ \mathbf{Z}'\mathbf{y}\end{bmatrix} - \begin{bmatrix}\tilde{\mathbf{a}}\\ \tilde{\mathbf{b}}\end{bmatrix}'\begin{bmatrix}0 & 0\\ 0 & k\mathbf{I}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{a}}\\ \tilde{\mathbf{b}}\end{bmatrix} \\
&\quad = \mathbf{y}'\mathbf{y} - R^*(\mathbf{y}, \mathbf{a}, \mathbf{b}) - k\tilde{\mathbf{b}}'\tilde{\mathbf{b}}.
\end{aligned}
$$

Cunningham and Henderson's method is:

(i) make a prior estimate of k,
(ii) estimate $\mathbf{a}$ by $\tilde{\mathbf{a}}$,
(iii) calculate $R^*(\mathbf{y}, \mathbf{a}, \mathbf{b})$ and $R^*(\mathbf{b} \mid \mathbf{y}, \mathbf{a})$,
(iv) estimate σ_e^2 , σ_b^2 from the mean squares,
(v) calculate k and iterate from stage (ii).

The expectations of the various sums of squares for a given value k can be deduced using

$$E[\mathbf{y}'\mathbf{A}\mathbf{y}] = \text{tr } \mathbf{A} \text{ var } (\mathbf{y}) + E(\mathbf{y}')\mathbf{A}E(\mathbf{y})$$

where tr is the trace operator. Then, as shown by Cunningham and Henderson [1968],

$$E[R(\mathbf{y}, \mathbf{a})] = \mathbf{a}'\mathbf{X}'\mathbf{X}\mathbf{a} + q'\sigma_e^2 + [\text{tr } \mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}]\sigma_b^2 \tag{2}$$

$$E(\mathbf{y}'\mathbf{y}) = \mathbf{a}'\mathbf{X}'\mathbf{X}\mathbf{a} + N\sigma_e^2 + N\sigma_b^2 \; ; \tag{3}$$

but

$$E[R^*(\mathbf{b} \mid \mathbf{y}, \mathbf{a})] = (\text{tr } \mathbf{Z}'\mathbf{S}\mathbf{Z}\mathbf{T})\sigma_e^2 + (\text{tr } \mathbf{Z}'\mathbf{S}\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{S}\mathbf{Z})\sigma_b^2 \tag{4}$$

and not

$$E[R^*(\mathbf{b} \mid \mathbf{y}, \mathbf{a})] = (p' - q')\sigma_e^2 + (\text{tr } \mathbf{T})\sigma_b^2 \tag{5}$$

as Cunningham and Henderson imply.

Now $R^*(\mathbf{b} \mid \mathbf{y}, \mathbf{a})$ is a function of k, $\mathbf{y}'\mathbf{y} - R(\mathbf{y}, \mathbf{a})$ is independent of k, and $\mathbf{y}'\mathbf{y} - R^*(\mathbf{y}, \mathbf{a}, \mathbf{b}) = \mathbf{y}'\mathbf{y} - R(\mathbf{y}, \mathbf{a}) - R^*(\mathbf{b} \mid \mathbf{y}, \mathbf{a})$ is a function of k. Cunningham and Henderson give $(N - p')\sigma_e^2$ as their expectation of $\mathbf{y}'\mathbf{y} - R^*(\mathbf{y}, \mathbf{a}, \mathbf{b})$ and as this is independent of k a contradiction is evident in their work.

When $k = 0$, equations (2), (3) and (4) give

$$E[R^*(\mathbf{b} \mid \mathbf{y}, \mathbf{a})] = E[R(\mathbf{b} \mid \mathbf{y}, \mathbf{a})] = (p' - q')\sigma_e^2 + (\text{tr } \mathbf{Z}'\mathbf{S}\mathbf{Z})\sigma_b^2 \tag{6}$$

and

$$E[\mathbf{y}'\mathbf{y} - R^*(\mathbf{y}, \mathbf{a}, \mathbf{b})] = E[\mathbf{y}'\mathbf{y} - R(\mathbf{y}, \mathbf{a}, \mathbf{b})] = (N - p')\sigma_e^2 \; . \tag{7}$$

Equations (6) and (7) agree with Cunningham and Henderson [1968] in the text although for their Table 1 to be correct

$$q' = \text{rank } (\mathbf{X}) - 1, \qquad r' = \text{rank } (\mathbf{Z}) - 1.$$

The evaluation of (4) can be long and tedious, especially if the number of random classes is large. The same solution can be reached in a much easier and more satisfying manner if we note that throughout the iteration, or at least after the first round, k is an estimate of σ_e^2/σ_b^2. Hence (4) can be expanded in the form

$$E[R^*(\mathbf{b} \mid \mathbf{y}, \mathbf{a})] = [\text{rank } (\mathbf{T}) - k \text{ tr } \mathbf{T}]\sigma_e^2 + [\text{tr } \mathbf{Z}'\mathbf{S}\mathbf{Z} - k \text{ rank } (\mathbf{T}) + k^2 \text{ tr } \mathbf{T}]\sigma_b^2 .$$

Substituting σ_e^2/σ_b^2 for k reduces this to

$$E[R^*(\mathbf{b} \mid \mathbf{y}, \mathbf{a})] = (\text{tr } \mathbf{Z}'\mathbf{S}\mathbf{Z})\sigma_b^2 . \tag{8}$$

Using (8) in conjunction with (2) and (3) we find

$$E[\mathbf{y}'\mathbf{y} - R^*(\mathbf{y}, \mathbf{a}, \mathbf{b})] = (N - q')\sigma_e^2 . \tag{9}$$

So σ_e^2 and σ_b^2 can be estimated using

$$\tilde{\sigma}_e^2 = [\mathbf{y}'\mathbf{y} - R^*(\mathbf{y}, \mathbf{b}) - R^*(\mathbf{a} \mid \mathbf{y}, \mathbf{b})]/(N - q'),$$

$$\tilde{\sigma}_b^2 = [R^*(\mathbf{y}, \mathbf{b}) + R^*(\mathbf{a} \mid \mathbf{y}, \mathbf{b}) - R^*(\mathbf{y}, \mathbf{a})]/c(p' - q'),$$

where c is defined in Table 1.

Using (8) and (9) as the basis of the iterative procedure avoids evaluating, for every round of the iteration, the complex expressions in (4). Estimates of σ_b^2 and σ_e^2 are available from (8) and (9) without the need to solve simultaneous equations. A satisfactory rate of convergence of estimates with this alternative method is dependent on k being a good estimate of σ_e^2/σ_b^2. Generally it will be better to use $k = 0$ with (6) and (7) to initiate the iteration procedure.

NUMERICAL EXAMPLE

Table 2 shows the results of applying the iterative procedure with formulae (8) and (9) to the illustrative example used by Cunningham and Henderson. The estimates stabilize after 5 iterations.

For the column headed 'Round 1', $k = 0$ was taken and the analysis proceeds as in Table 1. This gives new estimates of σ_e^2, σ_b^2, and k, in particular $k = 0.5721$. The column headed 'Round 2' is calculated using $k = 0.5721$ in (1) to give $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$. Then (8) and (9) give new estimates of σ_e^2, σ_b^2, and k. After 5 rounds, k and σ_e^2 have converged to values slightly smaller than those given in round 1 and σ_b^2 has converged to a value slightly larger. By contrast, Cunningham and Henderson's estimate of k continued increasing for 5 iterations. Indeed their formulae give

TABLE 2

ANALYSIS OF VARIANCE AND RESULTING ESTIMATES FOR FIVE ITERATIONS

Source of variance	Round 1	Round 2	Round 3	Round 4	Round 5
Total	437.0000	437.0000	437.0000	437.0000	437.0000
a elim. μ, b	15.1685	39.4080	38.8506	38.7486	38.7289
μ, b ign. a	386.5000	357.7526	358.4120	358.5330	358.5563
μ, a, b	401.6685	397.1606	397.2626	397.2816	397.2852
μ, a ign. b	350.5250	350.5250	350.5250	350.5250	350.5250
b elim. μ, a	51.1435	46.6356	46.7375	46.7566	46.7602
'Residual'	35.3315	39.8394	39.7374	39.7184	39.7148
$\bar{\sigma}_e^2$	2.5237	2.4900	2.4836	2.4824	2.4822
$\bar{\sigma}_b^2$	4.4111	4.4627	4.4725	4.4743	4.4747
k	0.5721	0.5579	0.5553	0.5548	0.5547
$\bar{a}_1 - \bar{a}_2$	−2.0225	−1.9313	−1.9333	−1.9337	−1.9338

$$\sigma_e^2 = 3.148, \quad \sigma_b^2 = 2.559, \quad k = 1.230 \quad \text{for round } 9,$$

and

$$\sigma_e^2 = 3.152, \quad \sigma_b^2 = 2.549, \quad k = 1.237 \quad \text{for round } 10,$$

showing that the trends apparent in their first five iterations were still continuing after 10 iterations. Their estimates differ from the corrected estimates, especially those of σ_b^2 and k.

DISCUSSION

In completely balanced cases, that is if blocks are of equal size and treatments are replicated equally within each block, the equations (2), (3), and (4) give the usual analysis of variance solution no matter what value of k is chosen.

Although the corrected iterative procedure leads to consistent estimation of σ_b^2 and σ_e^2, the estimates are not efficient. Indeed, the efficiency of the estimates after a convergent iteration may be lower than that of the estimates obtained at the end of round 1. If σ_b^2 is large relative to σ_e^2, estimation of σ_b^2 should give nearly the same weight to each block mean. For this, $k = 0$ and the estimation given by Table 1 will be the best approximation for the estimation of the variance components. On the other hand, if σ_b^2 is small, each observation should have nearly the same weight and the corrected Cunningham and Henderson method may then be successful. Indeed, the resultant 'residual' sum of squares when $\sigma_b^2 = 0$ is composed of the sum of squares due to blocks eliminating treatments and the residual sum of squares in the ordinary analysis of Table 1. This is equated to $(N - q')\sigma_e^2$, where $N - q'$ are the 'pooled' degrees of freedom.

ACKNOWLEDGEMENTS

I would like to thank Professor D. J. Finney and H. D. Patterson for their encouragement.

ESTIMATION ITERATIVE DES COMPOSANTES DE VARIANCE POUR DES DONNEES NON ORTHOGONALES

RESUME

Cunningham et Henderson [1968] ont présenté une procédure itérative pour l'estimation des constantes et des composantes de variances pour des données conformes au modèle mixte. On montre que cette méthode ne convient pas et des corrections sont suggérées. On montre que, pour un exemple, les estimations trouvées sont très différentes de celles de Cunningham et Henderson [1968]. La valeur pratique de cette méthode est discutée.

REFERENCES

Cunningham, E. P. and Henderson, C. R. [1968]. An iterative procedure for estimating fixed effects and variance components in mixed model situations. *Biometrics 24*, 13–25.

Eisenhart, C. [1947]. The assumptions underlying the analysis of variance. *Biometrics 3*, 1–21.

Nelder, J. A. [1968]. The combination of information in generally balanced designs. *J. R. Statist. Soc. B 30*, 303–11.

Received November 1968, *Revised June* 1969

2

Recovery of inter-block information when block sizes are unequal

by

H.D. Patterson and R. Thompson

Biometrika (1971), **58**, 3, *p.* 545
Printed in Great Britain

Recovery of inter-block information when block sizes are unequal

BY H. D. PATTERSON AND R. THOMPSON
University of Edinburgh

SUMMARY

A method is proposed for estimating intra-block and inter-block weights in the analysis of incomplete block designs with block sizes not necessarily equal. The method consists of maximizing the likelihood, not of all the data, but of a set of selected error contrasts. When block sizes are equal results are identical with those obtained by the method of Nelder (1968) for generally balanced designs. Although mainly concerned with incomplete block designs the paper also gives in outline an extension of the modified maximum likelihood procedure to designs with a more complicated block structure.

1. INTRODUCTION

In this paper we consider the estimation of weights to be used in the recovery of inter-block information in incomplete block designs with possibly unequal block sizes. The problem can also be thought of as one of estimating constants and components of variance from data arranged in a general two-way classification when the effects of one classification are regarded as fixed and the effects of the second classification are regarded as random.

Nelder (1968) described the efficient estimation of weights in generally balanced designs, in which the blocks are usually, although not always, of equal size. Lack of balance resulting from unequal block sizes is, however, common in some experimental work, for example in animal breeding experiments. The maximum likelihood procedure described by Hartley & Rao (1967) can be used but does not give the same estimates as Nelder's method in the balanced case.

As will be shown, the two methods in effect use the same weighted sums of squares of residuals but assign different expectations. In the maximum likelihood approach, expectations are taken over a conditional distribution with the treatment effects fixed at their estimated values. In contrast Nelder uses unconditional expectations.

The difference between the two methods is analogous to the well-known difference between two methods of estimating the variance σ^2 of a normal distribution, given a random sample of n values. Both methods use the same total sum of squares of deviations. But whereas one method equates the sum of squares to $(n-1)\sigma^2$ the other equates the sum of squares to $n\sigma^2$. The former method gives an unbiased estimate of σ^2; the latter maximizes the likelihood of the sample.

Another method for unbalanced designs has been described by Cunningham & Henderson (1968). When corrected as described by Thompson (1969) this method allows for errors in the estimation of treatment effects but in general the estimates are not efficient.

The method proposed in the present paper is a modified maximum likelihood procedure, more efficient than the Cunningham and Henderson method and giving the same results as Nelder's method in the analysis of balanced designs. The contrasts among yields are divided into two sets: (i) contrasts between treatment totals; and (ii) contrasts with zero

expectation, i.e. error contrasts. The method consists of maximizing the joint likelihood of all possible contrasts in set (ii). Contrasts in set (i) are excluded from the likelihood function on the grounds that, as long as treatment effects are regarded as unknown, fixed, as opposed to random, and without restraints, no contrast in set (i) can provide any information on error.

The modified maximum likelihood method was considered by Patterson (1964) in a components-of-variance problem arising in the analysis of rotation experiments. This paper was, however, primarily concerned with evaluating the efficiencies of simpler methods and gave no details beyond formulae for asymptotic variances.

2. THE MODEL

We suppose that the incomplete block design has t treatments and n units, plots, in b blocks of possibly unequal size and that the $n \times 1$ vector of yields, $\mathbf{y}$, can be represented by the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{X}$ is an $n \times t$ matrix of rank t determined by the allocation of treatments to units, $\boldsymbol{\alpha}$ is a $t \times 1$ vector of treatment parameters and $\boldsymbol{\epsilon}$ is a random variable normally distributed with mean zero and variance given by

$$\mathbf{V} = \mathbf{H}\sigma^2, \quad \mathbf{H} = \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}' + \mathbf{I}. \tag{2}$$

Further $\boldsymbol{\Gamma} = \gamma\mathbf{I}$, $\mathbf{Z}$ is an $n \times b$ matrix with elements Z_{ij} equal to 1 when unit i is in block j ($i = 1, \ldots, n$; $j = 1, \ldots, b$) and equal to 0 otherwise, and γ and σ^2 are unknown scalars. A model with more general $\mathbf{Z}$ and $\boldsymbol{\Gamma}$ is considered in § 10.

The problem is to estimate $\boldsymbol{\alpha}$, γ and σ^2. Sometimes reparameterization may be possible. Provided γ is not negative, $\gamma\sigma^2$ and σ^2 can be regarded as components of variance σ_b^2 and σ_e^2, say. When blocks sizes are all equal to k we may require to estimate functions

$$V_1 = (k\gamma + 1)\sigma^2, \quad V_2 = \sigma^2.$$

This is, in fact, what Nelder (1968) does.

The matrix $\mathbf{H}$ is essentially nonsingular. The inverse $\mathbf{H}^{-1}$ can be written in the form

$$\mathbf{H}^{-1} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \boldsymbol{\Gamma}^{-1})^{-1}\mathbf{Z}'. \tag{3}$$

Thus $\mathbf{H}^{-1}$ exists if $(\mathbf{Z}'\mathbf{Z} + \boldsymbol{\Gamma}^{-1})^{-1}$ exists; $\mathbf{Z}'\mathbf{Z}$ is diagonal with elements k_j, where k_j is the number of plots in block j. Clearly $\mathbf{Z}'\mathbf{Z} + \boldsymbol{\Gamma}^{-1}$ is singular if $\gamma = -k_j^{-1}$ for some j. But this would imply that the variance of the mean for block j is zero. A smaller value of γ implies a negative variance. We can, therefore, reasonably impose the condition that $\gamma > -1/k_{\max}$, where $k_{\max}$ is the largest number of plots in a single block.

Other conditions must also be satisfied for estimation of γ and σ^2 to be possible. These conditions will be considered in § 6.

Error model (2) also implies that the correlation between two units in the same block is independent of block size. This model is often used in animal experiments with blocks consisting of genetically related animals; usually the relationships within a block can be assumed to be independent of the number of animals in the block. In some other applications, for example field plot experiments, it might be more appropriate to specify a correlation that varies from block to block with the largest blocks showing the smallest correlations.

Estimates of $\boldsymbol{\alpha}$, γ and σ^2 will be denoted by $\hat{\boldsymbol{\alpha}}$, $\hat{\gamma}$ and $\hat{\sigma}^2$. We will also use the circumflex to denote functions of $\hat{\gamma}$, for example $\hat{\boldsymbol{\Gamma}} = \hat{\gamma}\mathbf{I}$ and $\hat{\mathbf{H}} = \mathbf{Z}\hat{\boldsymbol{\Gamma}}\mathbf{Z}' + \mathbf{I}$.

3. Modified maximum likelihood method

The logarithm of the likelihood function of $\mathbf{y}$ is given by

$$L = \text{const} - \tfrac{1}{2}\log|\mathbf{H}| - \tfrac{1}{2}n\log\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha}). \tag{4}$$

Hartley & Rao (1967) estimate $\boldsymbol{\alpha}$, γ and σ^2 by maximizing L. In the present paper we divide the data into two parts, with separate logarithmic likelihoods L' and L'', estimate γ and σ^2 by maximizing L' and estimate $\boldsymbol{\alpha}$ by maximizing L''.

The two parts can be represented by transformed yields $\mathbf{Sy}$ and $\mathbf{Qy}$ with the following properties.

(i) The matrix $\mathbf{S}$ is of rank $n-t$ and $\mathbf{Q}$ is a matrix of rank t.

(ii) The two parts are statistically independent, i.e. $\text{cov}(\mathbf{Sy}, \mathbf{Qy}) = 0$.
This condition is met if

$$\mathbf{SHQ}' = 0. \tag{5}$$

(iii) The matrix $\mathbf{S}$ is chosen so that

$$E(\mathbf{Sy}) = 0, \quad \text{i.e. } \mathbf{SX} = 0. \tag{6}$$

(iv) The matrix $\mathbf{QX}$ is of rank t, so that every linear function of the elements of $\mathbf{Qy}$ estimates a linear function of the elements of $\boldsymbol{\alpha}$.

It follows from (i) and (ii) that the likelihood of $\mathbf{y}$ is the product of the likelihoods of $\mathbf{Sy}$ and $\mathbf{Qy}$, i.e.

$$L = L' + L''. \tag{7}$$

Suitable matrices $\mathbf{S}$ and $\mathbf{Q}$ are given by

$$\mathbf{S} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}', \tag{8}$$

$$\mathbf{Q} = \mathbf{X}'\mathbf{H}^{-1}. \tag{9}$$

The matrix $\mathbf{S}$ is symmetric, idempotent, of rank $n-t$ and independent of γ. The elements of $\mathbf{Sy}$ are deviations of yields from treatment means. An estimate of γ is required for the transformation $\mathbf{Qy}$.

4. Estimation of γ and σ^2

First, we estimate γ and σ^2 by maximizing L', the logarithmic likelihood of $\mathbf{Sy}$. The variance matrix $\mathbf{SHS}\sigma^2$ is singular but suitable expressions for L' are available in terms of either a generalized inverse $(\mathbf{SHS})^{-g}$ or the latent roots of $\mathbf{SHS}$.

In deriving these expressions we use an $n \times (n-t)$ matrix $\mathbf{P}$ whose columns are orthogonal vectors of both $\mathbf{S}$ and $\mathbf{SHS}$. As $\mathbf{S}$ is idempotent and symmetric, it can be expressed in the form $\mathbf{AA}'$, where $\mathbf{A}$ is an $n \times (n-t)$ matrix such that $\mathbf{A}'\mathbf{A} = \mathbf{I}$. Now let $\mathbf{B}$ be an orthogonal matrix such that $\mathbf{B}'\mathbf{A}'\mathbf{HAB}$ is diagonal.

The required matrix $\mathbf{P}$ is given by $\mathbf{AB}$. It has the following properties: (i) $\mathbf{P}'\mathbf{P} = \mathbf{I}$, (ii) $\mathbf{PP}' = \mathbf{S}$ and (iii) $\mathbf{P}'\mathbf{HP} = \text{diag}(\xi_s)$, a diagonal matrix with elements $\xi_s (s = 1, \ldots, n-t)$. We note from (i) and (ii) that $\mathbf{P}'\mathbf{S} = \mathbf{P}'$. Hence $\mathbf{P}'\mathbf{y}$ can be derived from $\mathbf{Sy}$ by the transformation $\mathbf{P}'(\mathbf{Sy})$. Also $\mathbf{Sy}$ can be derived from $\mathbf{P}'\mathbf{y}$ by the transformation $\mathbf{P}(\mathbf{P}'\mathbf{y})$. It follows that the likelihood of $\mathbf{Sy}$ is also the likelihood of $\mathbf{P}'\mathbf{y}$.

Property (iii) shows that

$$\mathbf{P}'\mathbf{HS} = \text{diag}\,(\xi_s)\,\mathbf{P}', \tag{10}$$

i.e. the ξ_s are the nonzero latent roots of **HS**. We can now write **SHS** in the spectral form

$$\mathbf{SHS} = \mathbf{P}\,\text{diag}\,(\xi_s)\,\mathbf{P}' \tag{11}$$

and define a generalized inverse $(\mathbf{SHS})^{-g}$ such that

$$(\mathbf{SHS})^{-g} = \mathbf{P}\,\text{diag}\,(\xi_s^{-1})\,\mathbf{P}'. \tag{12}$$

The product of $(\mathbf{SHS})^{-g}$ with **SHS** is **S**. It is also worth noting that $(\mathbf{SHS})^{-g}$ is unchanged on multiplication by **S**.

The elements of $\mathbf{P}'\mathbf{y}$ are contrasts $u_s (s = 1, \ldots, n-t)$ with variances $\xi_s\sigma^2$. Hence the required log likelihood function is given by

$$L' = \text{const} - \tfrac{1}{2}\sum_s \log \xi_s - \tfrac{1}{2}(n-t)\log\sigma^2 - R/(2\sigma^2), \tag{13}$$

where R is the weighted sum of squares of the u_s given by

$$R = \Sigma\,(u_s^2/\xi_s) = \mathbf{y}'(\mathbf{SHS})^{-g}\,\mathbf{y}. \tag{14}$$

The estimates $\hat{\gamma}$ and $\hat{\sigma}^2$ maximizing L' are obtained by solving the equations

$$\frac{\partial L'}{\partial \gamma} = -\tfrac{1}{2}E + \frac{B}{2\sigma^2} = 0, \tag{15}$$

$$\frac{\partial L'}{\partial \sigma^2} = -\frac{1}{2}\frac{n-t}{\sigma^2} + \frac{R}{2\sigma^4} = 0, \tag{16}$$

where

$$B = \sum_s u_s^2 d_s/\xi_s^2 = \mathbf{y}'(\mathbf{SHS})^{-g}\,\mathbf{SDS}(\mathbf{SHS})^{-g}\,\mathbf{y}, \tag{17}$$

$$E = \sum_s (d_s/\xi_s) = \text{tr}\,\{(\mathbf{SHS})^{-g}\,\mathbf{SDS}\}. \tag{18}$$

Further d_s denotes $\partial\xi_s/\partial\gamma$ and **D** denotes $\partial\mathbf{H}/\partial\gamma$.

The expected values of $\partial L'/\partial\gamma$ and $\partial L'/\partial\sigma^2$ for fixed γ are both zero. Solution of equations (15) and (16) consists therefore of equating B and R to their expectations in the conditional distribution with γ fixed.

The information matrix is

$$\frac{1}{2}\begin{bmatrix} f_{11} & f_{12}/\sigma^2 \\ f_{12}/\sigma^2 & f_{22}/\sigma^4 \end{bmatrix},$$

where

$$f_{11} = \sum_s (d_s/\xi_s)^2 = \text{tr}\,\{(\mathbf{SHS})^{-g}\,\mathbf{SDS}\}^2, \tag{19}$$

$$f_{12} = E, \tag{20}$$

$$f_{22} = n-t. \tag{21}$$

So far the results are appropriate when **H** is any symmetric $n \times n$ matrix of rank n such that $\text{rank}(\mathbf{SHS}) = \text{rank}(\mathbf{S}) = n-t$. For practical purposes, however, and for comparison with other methods of estimation, we require simpler expressions for B, R, E and the elements of the information matrix. When **H** takes the particular form specified in (2) suitable expressions can be obtained by substituting the appropriate differentials **D** and d_s and a simplified form of $(\mathbf{SHS})^{-g}$.

The differential $\mathbf{D}$ is simply $\mathbf{ZZ'}$ and the d_s are given by the latent roots of $\mathbf{ZZ'S}$, or $\mathbf{SZZ'S}$. The latter can be demonstrated by substituting for $\mathbf{H}$ in equation (11) and rearranging to give

$$\mathbf{SZZ'S} = \mathbf{P}\,\mathrm{diag}\,(\lambda_s)\,\mathbf{P'}, \tag{22}$$

where the λ_s are such that

$$\xi_s = \lambda_s\gamma + 1. \tag{23}$$

Hence $d_s = \lambda_s$.

We now express $(\mathbf{SHS})^{-g}$ in terms of a $b \times b$ matrix. First, we note that the factors of $\mathbf{Z'SZ\Gamma Z'S}$ include both the $b \times b$ matrix $\mathbf{Z'SZ}$ and the $n \times n$ matrix $\mathbf{SZ\Gamma Z'S} = \mathbf{S(H-I)S}$. Hence

$$(\mathbf{Z'SZ})\,\mathbf{\Gamma Z'S} + \mathbf{Z'S} = \mathbf{Z'S(SHS)},$$

so that

$$\mathbf{\Gamma Z'}(\mathbf{SHS})^{-g} = \mathbf{W}^{-1}\mathbf{Z'S}, \tag{24}$$

where

$$\mathbf{W} = \mathbf{Z'SZ} + \mathbf{\Gamma}^{-1}. \tag{25}$$

Premultiplying both sides of (24) by $\mathbf{SZ}$, we obtain

$$(\mathbf{SHS} - \mathbf{S})\,(\mathbf{SHS})^{-g} = \mathbf{SZW}^{-1}\mathbf{Z'S},$$

so that

$$(\mathbf{SHS})^{-g} = \mathbf{S} - \mathbf{SZW}^{-1}\mathbf{Z'S}. \tag{26}$$

The quantities B, R, E and f_{11} can now be expressed in terms of $\mathbf{W}$, a vector

$$\boldsymbol{\beta} = \mathbf{\Gamma Z'}(\mathbf{SHS})^{-g}\,\mathbf{y} = \mathbf{W}^{-1}\mathbf{Z'Sy} \tag{27}$$

and a matrix

$$\mathbf{U} = \mathbf{Z'}(\mathbf{SHS})^{-g}\mathbf{Z} = \mathbf{\Gamma}^{-1} - \mathbf{\Gamma}^{-1}\mathbf{W}^{-1}\mathbf{\Gamma}^{-1}. \tag{28}$$

The required expressions for B and R can be obtained by direct substitution. They are

$$B = \boldsymbol{\beta'\beta}/\gamma^2, \tag{29}$$

$$R = \mathbf{y'Sy} - \mathbf{y'SZ}\boldsymbol{\beta}. \tag{30}$$

Also, since $\mathrm{tr}\,\{(\mathbf{SHS})^{-g}\mathbf{ZZ'S}\}$ is equivalent to $\mathrm{tr}\,\{\mathbf{Z'}(\mathbf{SHS})^{-g}\mathbf{Z}\}$, E is given by

$$E = f_{12} = \mathrm{tr}\,(\mathbf{U}). \tag{31}$$

Similarly,

$$f_{11} = \mathrm{tr}\,(\mathbf{U}^2). \tag{32}$$

In practice, of course, we use $\hat{\mathbf{H}}$, $\hat{B}$, $\hat{\boldsymbol{\beta}}$, $\hat{R}$, etc., given by the expressions for $\mathbf{H}$, B, $\boldsymbol{\beta}$, R, etc., but with γ replaced by $\hat{\gamma}$.

These results depend on $\mathbf{W}$ being nonsingular; in fact $\mathbf{W}$ is always nonsingular under the conditions imposed on $\mathbf{H}$ in §2 as the largest λ_s is necessarily not larger than $k_{\max}$.

5. Estimation of $\boldsymbol{\alpha}$

The estimate of $\boldsymbol{\alpha}$ is obtained by maximizing

$$L'' = \mathrm{const} - \tfrac{1}{2}\log|\mathbf{X'H}^{-1}\mathbf{X}| - \tfrac{1}{2}t\log\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}\mathbf{X}(\mathbf{X'H}^{-1}\mathbf{X})^{-1}\mathbf{X'H}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha}). \tag{33}$$

The estimate is given by

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X'}\hat{\mathbf{H}}^{-1}\mathbf{X})^{-1}\mathbf{X'}\hat{\mathbf{H}}^{-1}\mathbf{y}. \tag{34}$$

An equivalent procedure is to minimize the weighted sum of squares

$$(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})'\,\hat{\mathbf{H}}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})$$

with fixed $\hat{\gamma}$.

6. Practical solution of equations (15) and (16)

Equations (15) and (16) can be solved by Fisher's iterative method. We start with a preliminary estimate, $\hat{\gamma}_0$ say, of γ. Substitute this estimate for γ in (25) to give $\hat{\mathbf{W}}$, and hence determine $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{U}}$, $\hat{B}$, $\hat{R}$, $\hat{f}_{12}$ and $\hat{f}_{11}$ equations from (27) to (32). Then an approximation to $\hat{\sigma}^2$ is given by

$$\hat{\sigma}^2 = \hat{f}^{12}\hat{B}+\hat{f}^{22}\hat{R}, \tag{35}$$

and a closer approximation to $\hat{\gamma}$ is given by

$$\hat{\gamma} = \hat{\gamma}_0+(\hat{f}^{11}\hat{B}+\hat{f}^{12}\hat{R})/\hat{\sigma}^2. \tag{36}$$

In these expressions $\hat{f}^{11}$, $\hat{f}^{12}$ and $\hat{f}^{22}$ are the elements of the inverse matrix $\hat{\mathbf{F}}^{-1}$, where

$$\hat{\mathbf{F}} = \begin{bmatrix} \hat{f}_{11} & \hat{f}_{12} \\ \hat{f}_{12} & \hat{f}_{22} \end{bmatrix}.$$

If $\hat{\gamma}_0$ is small it may be more accurate numerically not to calculate $\hat{\boldsymbol{\beta}}$ explicitly, but to use instead

$$\hat{\boldsymbol{\beta}}/\hat{\gamma} = (\hat{\gamma}_0\mathbf{Z}'\mathbf{S}\mathbf{Z}+\mathbf{I})^{-1}\mathbf{Z}'\mathbf{S}\mathbf{y}.$$

A preliminary estimate of σ^2 is not required. Also $\boldsymbol{\alpha}$ need not be estimated until $\hat{\gamma}$ has been determined. A convenient expression for $\hat{\boldsymbol{\alpha}}$ in terms of $\hat{\gamma}$ can be obtained by noting that $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ in (34) and (27) satisfy the equations

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\alpha}}+\mathbf{X}'\mathbf{Z}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}, \tag{37}$$

$$\mathbf{Z}'\mathbf{X}\hat{\boldsymbol{\alpha}}+(\mathbf{Z}'\mathbf{Z}+\hat{\boldsymbol{\Gamma}}^{-1})\,\hat{\boldsymbol{\beta}} = \mathbf{Z}'\mathbf{y}. \tag{38}$$

Hence

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y}-\mathbf{Z}\hat{\boldsymbol{\beta}}). \tag{39}$$

Equations (37) and (38) were suggested by Henderson (Henderson, Kempthorne, Searle & von Krosigk, 1959) as the basis of a practical procedure for estimating $\boldsymbol{\alpha}$ when γ is known.

When b is large an alternative procedure may be preferable, requiring inversion of a $t\times t$ matrix and a diagonal $b\times b$ matrix instead of the nondiagonal $\mathbf{W}$.

Equation (34) is used first to estimate $\boldsymbol{\alpha}$. Equation (3) provides a convenient expression for $\hat{\mathbf{H}}^{-1}$, involving inversion of a diagonal matrix only. Then

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z}+\hat{\boldsymbol{\Gamma}}^{-1})^{-1}\mathbf{Z}'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}}), \tag{40}$$

a rearrangement of equation (38). The value of $\hat{B}$ is calculated as before and $\hat{R}$ is given by

$$\hat{R} = \mathbf{y}'\mathbf{y}-\mathbf{y}'\mathbf{X}\hat{\boldsymbol{\alpha}}-\mathbf{y}'\mathbf{Z}\hat{\boldsymbol{\beta}}. \tag{41}$$

The matrix $\hat{\mathbf{W}}^{-1}$, required in the calculation of $\hat{f}_{11}$ and $\hat{f}_{12}$, can be conveniently expressed in terms of the matrices $(\mathbf{Z}'\mathbf{Z}+\hat{\boldsymbol{\Gamma}}^{-1})^{-1}$ and $(\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}$ already used in the estimation of $\boldsymbol{\alpha}$. The expression is

$$\hat{\mathbf{W}}^{-1} = (\mathbf{Z}'\mathbf{Z}+\hat{\boldsymbol{\Gamma}}^{-1})^{-1}+(\mathbf{Z}'\mathbf{Z}+\boldsymbol{\Gamma}^{-1})^{-1}\mathbf{Z}'\mathbf{X}(\mathbf{X}'\hat{\mathbf{H}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z}+\hat{\boldsymbol{\Gamma}}^{-1})^{-1}. \tag{42}$$

Solution of equations (15) and (16) breaks down if $\mathbf{F}$ is singular. We distinguish three cases. When $f_{22} = 0$, i.e. $n = t$, neither γ nor σ^2 can be estimated as the treatment contrasts account for all $n-1$ degrees of freedom. When $f_{11} = 0$ but $f_{22} \neq 0$, σ^2 can be estimated but not γ. This situation arises when some treatment comparison is totally confounded with every block comparison. Each column of $\mathbf{Z}$ is then given by a linear combination of the columns of $\mathbf{X}$ so that $\mathbf{SZ} = 0$ and $\mathbf{W}^{-1} = \boldsymbol{\Gamma}$, and hence $\mathbf{U} = 0$. The matrix $\mathbf{F}$ is also singular when treatment contrasts account for all intra-block degrees of freedom but not for all inter-block degrees of freedom. In this case $f_{11} = (n-t)/(\gamma+1)^2$, $f_{12} = (n-t)/(\gamma+1); (\gamma+1)\sigma^2$ can be estimated but not the individual γ and σ^2.

A further complication is that $\mathbf{X}'\mathbf{X}$ may be singular This can be dealt with by replacing $(\mathbf{X}'\mathbf{X})^{-1}$ and $(\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}$ in (8), (34), (39) and (42) by generalized inverses and making t the rank of $\mathbf{X}$ instead of the number of treatments.

7. Relationship with other methods

We now consider the relationship between I, the method described by Hartley & Rao (1967), II, the method proposed by Cunningham & Henderson (1968) and later modified by Thompson (1969) and III, the method of the present paper. These three methods give different estimates of γ but use essentially the same methods of estimating $\boldsymbol{\alpha}$ and σ^2 once $\hat{\gamma}$ has been determined.

Hartley & Rao (1967) show that the unconditional maximum likelihood estimates are obtained by solving the following equations for $\boldsymbol{\alpha}$, σ^2 and γ:

$$\frac{1}{\sigma^2}(\mathbf{X}'\mathbf{H}^{-1}\mathbf{y} - \mathbf{X}'\mathbf{H}^{-1}\mathbf{X}\boldsymbol{\alpha}) = 0, \tag{43}$$

$$-\tfrac{1}{2}\operatorname{tr}(\mathbf{H}^{-1}\mathbf{Z}\mathbf{Z}') + \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}\mathbf{Z}\mathbf{Z}'\mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}) = 0, \tag{44}$$

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}) = 0. \tag{45}$$

The numerical terms $(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}\mathbf{Z}\mathbf{Z}'\mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})$ and $(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})$ are equivalent to B and R. This can be demonstrated as follows. Consider the expressions for L, L' and L'' given in (4), (13) and (33). As $L = L' + L''$, we have

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}) = R + (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}). \tag{46}$$

Substitution of the solution for $\boldsymbol{\alpha}$ given in § 5, or the solution given by equation (43), shows that the second term on the right hand side of (46) is zero. Hence

$$R = (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}). \tag{47}$$

Differentiation of both sides of equation (47) with respect to γ gives

$$B = (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}\mathbf{Z}\mathbf{Z}'\mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}). \tag{48}$$

Thus the Hartley–Rao method consists of equating B and R to their expected values in the conditional distribution with both γ and $\boldsymbol{\alpha}$ fixed. It differs from method III in that in the latter the expected values take account of errors in the estimation of $\boldsymbol{\alpha}$.

Method II differs from method III in that the sum of squares $\boldsymbol{\beta}'\mathbf{Z}'\mathbf{S}\mathbf{y}$ is used instead of B.

Thus $\boldsymbol{\beta}'\mathbf{Z}'\mathbf{S}\mathbf{y}$ and R are equated to their expected values in the conditional distribution with fixed γ. As in method III the estimates of γ and σ^2 are wholly derived from $\mathbf{S}\mathbf{y}$. Method II does not, however, maximize the likelihood of $\mathbf{S}\mathbf{y}$ except in special cases when $\boldsymbol{\beta}$ is proportional to $\mathbf{Z}'\mathbf{S}\mathbf{y}$, for example in complete block designs and in symmetric balanced incomplete block designs.

8. EQUAL BLOCK SIZES

Nelder (1968) proposed a method for estimating stratum variances in a general class of balanced designs. We now show that application of this method to incomplete block designs with blocks of equal size k gives the same results as method III.

The method consists of equating the sums of squares of residuals,

$$(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}})'\left(\frac{\mathbf{Z}\mathbf{Z}'}{k}-\frac{\mathbf{1}\mathbf{1}'}{n}\right)(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}}), \tag{49}$$

$$(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}})'\left(\mathbf{I}-\frac{\mathbf{Z}\mathbf{Z}'}{k}\right)(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}}) \tag{50}$$

to their expectations, where $\mathbf{1}$ is a unit vector.

Equations (48) and (47) show that $\hat{B}$ and $\hat{R}$, the quantities used in method III, are also sums of squares of residuals. When block sizes are all equal to k the expressions for $\hat{B}$ and $\hat{R}$ simplify to

$$\hat{B} = (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}})'\,\mathbf{Z}\mathbf{Z}'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}})/(\hat{\gamma}k+1)^2,$$

$$\hat{R} = (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}})'\,(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}})-\hat{\gamma}\hat{B};$$

$\mathbf{1}'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}})$, the sum of residuals, is zero. Equating $\hat{B}$ and $\hat{R}$ to their expectations is therefore equivalent to equating the expressions (49) and (50) to their expectations, i.e. the method proposed in the present paper gives the same results as Nelder's (1968) method.

Table 1. *Data used by Cunningham & Henderson* (1968)

	Treatment				Totals
	1		2		
Block 1	3, 2,	5	2, 3,	5	10
2	2, 3, 5, 6, 7,	23	8, 8, 9	25	48
3	3	3	4, 4, 3, 2, 5	18	21
Totals		31		48	79

Table 2. *Estimation of* σ^2, γ *from the data of Table* 1. *Initial estimate of* $\gamma = 1$

	Cycle			
	1	2	3	4
B	6·3194	2·7584	2·8537	2·8505
R	42·7221	40·2145	40·2981	40·2952
$\hat{f}^{11}$	0·8155	1·8424	1·7817	1·7836
$10\hat{f}^{12}$	−0·8530	−1·2824	−1·2611	−1·2618
$10\hat{f}^{32}$	0·7142	0·7143	0·7143	0·7143
$\hat{\sigma}^2$	2·5123	2·5186	2·5185	2·5185
$\hat{\gamma}$	1·6006	1·5708	1·5718	1·5718

9. Example

We have applied the new method, method III, to the data of the example discussed by Cunningham & Henderson (1968) and Thompson (1969). The data are in Table 1. Table 2 gives the results of four cycles of the iterative procedure, starting with $\hat{\gamma}_0 = 1$. Convergence is rapid. Table 3 compares the results with those given by methods I and II and by the analysis of variance method described by Cunningham & Henderson (1968).

Table 3. *Comparison of estimates of* σ^2 *and* γ

Method	Estimate of σ^2	Estimate of γ
Analysis of variance	2·5237	1·7479
Method I	2·3518	1·0652
Method II	2·4822	1·8028
Method III	2·5185	1·5718

10. More general results

Hartley & Rao (1967) extended the unconditional maximum likelihood method to a general class of designs with c block factors. The class includes for example split-plot designs and row-and-column designs. With these designs the model given by equation (1) still applies but the variance matrix of $\boldsymbol{\epsilon}$ is now $\mathbf{H}\sigma^2$, where

$$\mathbf{H} = \left(\mathbf{I} + \sum_{p=1}^{c} \mathbf{Z}_p \mathbf{Z}_p' \gamma_p\right). \tag{51}$$

We have to estimate $c+1$ parameters, σ^2 and γ_p $(p = 1, \ldots, c)$.

Each plot is at exactly one level of each block factor. For example, if a design is arranged in rows and columns each plot is in exactly one row and one column. Element (i,j) of $\mathbf{Z}_p$ is 1 when plot i is at level j of block factor p $(j = 1, \ldots, b_p; p = 1, \ldots, c)$; otherwise the element is 0. The matrices $\mathbf{Z}_p'\mathbf{Z}_p$ are diagonal.

The modified maximum likelihood method of the present paper can also be applied to this more general class of designs. We again divide the data into two parts with logarithmic likelihoods L' and L'', estimate γ_p and σ^2 by maximizing L', and estimate $\boldsymbol{\alpha}$ by maximizing L''.

With suitable redefinition of $\boldsymbol{\Gamma}$ and $\mathbf{Z}$, results closely follow those already given for the simpler model. Only the main modifications will be described here.

We redefine $\boldsymbol{\Gamma}$ as a diagonal $b \times b$ matrix, i.e. $\boldsymbol{\Gamma} = \text{diag}\,(\gamma_p \mathbf{I}_p)$ $(p = 1, \ldots, c)$, where $b = \Sigma b_p$ and $\mathbf{I}_p$ is the $b_p \times b_p$ identity matrix; $\mathbf{Z}$ is the partitioned matrix $(\mathbf{Z}_1 \vdots \ldots \vdots \mathbf{Z}_c)$. With these definitions $\mathbf{H}$ can again be written in the form $\mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}' + \mathbf{I}$. The vector $\boldsymbol{\beta}$ and the matrix $\mathbf{U}$ defined by (27) and (28) are also partitioned. The $b_p \times 1$ subvectors of $\boldsymbol{\beta}$ will be denoted by $\boldsymbol{\beta}_p$ and the $b_p \times b_q$ submatrices of $\mathbf{U}$ will be denoted by $\mathbf{U}_{pq}$ $(p, q = 1, \ldots, c)$.

The estimates $\hat{\gamma}_p$ and $\hat{\sigma}^2$ are obtained by solving the c equations

$$-\tfrac{1}{2}E_p + \frac{1}{2\sigma^2} B_p = 0 \quad (p = 1, \ldots, c),$$

together with equation (16), where

$$B_p = \mathbf{y}'(\mathbf{SHS})^{-g} \mathbf{Z}_p \mathbf{Z}_p' (\mathbf{SHS})^{-g} \mathbf{y},$$

$$E_p = \text{tr}\,\{(\mathbf{SHS})^{-g} \mathbf{Z}_p \mathbf{Z}_p' \mathbf{S}\}.$$

Working expressions for B_p and E_p are

$$B_p = \boldsymbol{\beta}_p'\boldsymbol{\beta}_p/\gamma_p^2, \quad E_p = \mathrm{tr}\,(\mathbf{U}_{pp}).$$

The information matrix is

$$\tfrac{1}{2}\begin{bmatrix} \mathrm{tr}\,(\mathbf{U}_{11}^2) & \mathrm{tr}\,(\mathbf{U}_{12}\mathbf{U}_{21}) & \dots & \mathrm{tr}\,(\mathbf{U}_{1c}\mathbf{U}_{c1}) & \mathrm{tr}\,(\mathbf{U}_{11})/\sigma^2 \\ \mathrm{tr}\,(\mathbf{U}_{12}\mathbf{U}_{21}) & \mathrm{tr}\,(\mathbf{U}_{22}^2) & \dots & \mathrm{tr}\,(\mathbf{U}_{2c}\mathbf{U}_{c2}) & \mathrm{tr}\,(\mathbf{U}_{22})/\sigma^2 \\ . & . & \dots & . & . \\ \mathrm{tr}\,(\mathbf{U}_{1c}\mathbf{U}_{c1}) & \mathrm{tr}\,(\mathbf{U}_{2c}\mathbf{U}_{c2}) & \dots & \mathrm{tr}\,(\mathbf{U}_{cc})^2 & \mathrm{tr}\,(\mathbf{U}_{cc})/\sigma^2 \\ \mathrm{tr}\,(\mathbf{U}_{11})/\sigma^2 & \mathrm{tr}\,(\mathbf{U}_{22})/\sigma^2 & \dots & \mathrm{tr}\,(\mathbf{U}_{cc})/\sigma^2 & (n-t)/\sigma^4 \end{bmatrix}.$$

Generalization of the iterative solution described in § 6 presents little difficulty. Singularities in the information matrix again indicate that one or more parameters cannot be estimated. For example, we cannot estimate γ_q when $\mathbf{U}_{qq} = 0$.

Further generality can be achieved by defining $\mathbf{Z}_p\mathbf{Z}_p'$ as any real symmetric matrix with $\mathbf{Z}_p'\mathbf{Z}_p$ not necessarily diagonal. We can also deal with variance matrices of the form $\mathbf{V} = (\mathbf{J} + \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}')\sigma^2$, where $\mathbf{J}$ is symmetric. When $\mathbf{J}$ is idempotent and such that $\mathbf{JZ} = \mathbf{Z}$, $\mathbf{JX} = \mathbf{X}$ and $\mathbf{J} \neq \mathbf{I}$, generalized inverses are required in place of $\mathbf{H}^{-1}$, $(\mathbf{X}'\mathbf{X})^{-1}$ and $(\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}$. When $\mathbf{J}$ is not idempotent a preliminary transformation $\mathbf{Ty}$ can be used with $\mathbf{T}$ such that $\mathbf{TJT}'$ is idempotent.

REFERENCES

CUNNINGHAM, E. P. & HENDERSON, C. R. (1968). An iterative procedure for estimating fixed effects and variance components in mixed model situations. *Biometrics* **24**, 13–25.

HARTLEY, H. O. & RAO, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* **54**, 93–108.

HENDERSON, C. R., KEMPTHORNE, O., SEARLE, S. R. & VON KROSIGK, C. M. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics* **15**, 192–218.

NELDER, J. A. (1968). The combination of information in generally balanced designs. *J. R. Statist. Soc.* B **30**, 303–11.

PATTERSON, H. D. (1964). Theory of cyclic rotation experiments (with discussion). *J. R. Statist. Soc.* B **26**, 1–45.

THOMPSON, R. (1969). Iterative estimation of variance components for non-orthogonal data. *Biometrics* **25**, 767–73.

[*Received March* 1971. *Revised June* 1971]

Some key words: Analysis of incomplete block designs; Components of variance; Designs with unequal block sizes; Recovery of inter-block information; Maximum likelihood.

3

Maximum likelihood estimation of components of variance

by

H.D. Patterson and R. Thompson

THE BIOMETRIC SOCIETY

Proceedings of the 8th International Biometric CONFERENCE

August 25-30
1974
Constanța
Romania

The conference was organized by:

THE NATIONAL GROUP OF BIOMETRY OF THE ACADEMY OF THE SOCIALIST REPUBLIC OF ROMANIA

THE CENTRE OF MATHEMATICAL STATISTICS OF THE MINISTRY OF EDUCATION OF ROMANIA

Edited by: **L.C.A CORSTEN**
TIBERIU POSTELNICU

REPRINT

EDITURA ACADEMIEI REPUBLICII SOCIALISTE ROMÂNIA

MAXIMUM LIKELIHOOD ESTIMATION OF COMPONENTS OF VARIANCE

H. D. PATTERSON and **R. THOMPSON**

A.R.C. Unit of Statistics, Edinburgh University

Estimation of the variance parameters of a mixed model by maximising the likelihood of the observed values suffers from what we regard as a serious defect, namely that its application to the analysis of highly organised data is inconsistent with long-established practice. A simple modification, maximising the likelihood of contrasts with expectation zero rather than the complete likelihood removes this defect. We are encouraged to note that the modified method gives results that are in line with Rao's MINQUE, and the method devised by Nelder (1968) to meet the special requirements of experimental design.

1. INTRODUCTION

This paper is concerned with the well-known problem of estimating variance parameters in the mixed model. The model is specified by the following equations :

$$E(Y) = X\beta, \tag{1.1}$$

$$D(Y) = V = V_1\theta_1 + V_2\theta_2 + \cdots + V_m\theta_m. \tag{1.2}$$

The first equation gives the expected value of a vector Y of observations in terms of a known design matrix X and a vector of unknown parameters β. The second equation gives the dispersion matrix of Y. Each element of the dispersion matrix is a linear function of m unknown variance parameters $\theta_1, \theta_2 \cdots \theta_m$. The matrices $V, V_1, V_2, \cdots V_m$, are symmetric but may be singular.

The elements of β are not functionally related to the variance parameters but may be subject to constraints resulting from singularities in V. Rao (1973, Section 4.i.2) has shown how these constraints arise. If q is a vector in the null space of V then the variance of the contrast q^TY is zero i.e. q^TY is constant. If also q is in the null space of X^T the constant is zero. But if q is not in the null space of X^T then the elements of β are subject to the following constraint

$$k^T\beta = \text{constant}, \quad k^T = q^TX. \tag{1.3}$$

No other constraints are imposed on β in the present paper and there are no constraints on the variance parameters.

The problem is to obtain estimates of the variance parameters. Given these we estimate β by the method described by Rao and Mitra (1971) for V known apart from a proportionality factor.

Throughout we use the mathematical terminology of Rao (1973) In particular $M(A)$ denotes the column space of A. The statements $M(A \subset M(B)$, $a \in M(B)$ mean that there is a matrix C and a vector c such tha $A = BC$, $a = Bc$. Also A^- is a generalised inverse of A and A^+ is the Moore Penrose inverse.

2. SPECIAL CASE : ORTHOGONAL DESIGNS

Of course we know how to solve the problem in certain special case and in particular in the analysis of experiments with orthogonal block and treatment structure. As an example we consider a split-plot design with three blocks, eight whole plots per block and four sub-plots per whole plot. The eight levels of one treatment factor (A) are allocated to whole plots and the four levels of a second treatment factor (B) to sub-plots. In this case, following Nelder (1965a), we give the dispersion matrix the form

$$V = C_0\theta_0 + C_1\theta_2 + C_1\theta_2 + C_3\theta_3$$

with C_0, C_1, C_2, C_3 idempotent orthogonal and summing to I. These matrices partition the total sum of squares into components Y^TC_0Y, Y^TC_1Y, Y^TC_2Y, Y^TC_3Y with degrees of freedom as shown in Table 1. The degrees of freedom in the whole-plot and sub-plot strata are further subdivided into treatment and error components. The parameter θ_1 is then estimated by the blocks mean square, θ_2 by the whole-plot error mean square and θ_3 by the sub-plot error mean square.

This method is well tried and practised. Through his teaching G.W. Snedecor, to whose memory this session is dedicated, contributed in no small measure to its understanding and almost universal acceptance.

Problems arise, however, when there are departures from orthogonality in block structure or treatment structure or both.
Analysis-of-variance methods can still be used in which mean squares are equated to expectation. Sometimes these methods can be useful but in general are neither efficient nor unique.

More general methods are, therefore, required. We consider some in the course of this paper. In looking at these methods we will have in mind desirable properties of efficiency and unbiasedness and other statistical criteria. But, whatever principle is invoked, we seek a general method that reduces to the analysis of variance in the completely orthogonal case and so at least gives us continuity of practice.

Table 1

Analysis of variance of results of split-plot experiment

	d.f.	d.f.	S.S.	M.S.	estimates
Block stratum:	2		Y^TC_1Y		θ_1
Whole-plot stratum:	21		Y^TC_2Y		
Factor A		7			
Error		14			θ_2
Sub-plot stratum:	72		Y^TC_3Y		
Factor B		3			
$A \times B$		21			
Error		48			θ_3

3. MORE GENERAL METHODS

A natural extension of the analysis-of-variance method was proposed by Nelder (1968) for application to a wide class of generally balanced designs. These designs still have orthogonal block structure so that the total sum of squares can be partitioned into components Y^TC_0Y, Y^TC_1Y etc. where C_0, C_1 etc., are orthogonal and idempotent. But the treatment structure is no longer orthogonal. Nelder's method consists of estimating treatment parameters as efficiently as possible, using weighted least squares when appropriate, calculating the vector of residuals $R\underset{\sim}{Y}$ and equating the sums of squares $Y^TR^TC_iR\underset{\sim}{Y}$ to expectation.

Still greater generality is provided by Rao's (1970) method of minimum norm quadratic unbiased estimation (MINQUE). The estimators are of the form

$$\hat{\theta}_i = Y^TA_iY. \tag{3.1}$$

The matrix A_i is chosen to minimise the Euclidean norm (sum of squares of all elements) of A_iU where U is a p.d. matrix given by

$$U = \alpha_1^2V_1 + \alpha_2^2V_2 + \cdots + \alpha_m^2V_m. \tag{3.2}$$

The minimisation is subject to the condition

$$E(Y^TA_iY) = \theta_i. \tag{3.3}$$

which ensures unbiased estimation of θ_i.

In general the estimates of θ_i depend on the choice of α_i. Rao (1973) recommended that when possible α_i^2 should be chosen approximately proportional to the θ_i. If there is no prior information on θ_i the α_i may be taken to be equal.

The introduction of minimum norm quadratic estimation has proved valuable and stimulated a great deal of research but the method depends on a measure that is not accepted or understood by all statisticians. Moreover, there is a danger of misapplication through unwise or unfortunate choice of the scale of the variance parameters.

It is, however, possible to justify the use of MINQUE by appeals to more conventional statistical principles such as minimum variance, least squares, maximum likelihood. Thus, Mitra (1971) has pointed out that MINQUE estimators have local minimum variance when $\underset{\sim}{Y}$ is n-normal.

In the present paper we are concerned primarily with maximum likelihood estimation. Hartley and J.N.K. Rao (1967) described the procedure to be followed when Y is n-normal, V is positive definite and the likelihood of the complete vector Y is to be maximised.

Unfortunately, this method gives results that are not consistent with long-established practices in the analysis of completely orthogonal data. In the split-plot example of section 2 the maximum likelihood method

gives estimates of θ_1, θ_2 and θ_3 only two-thirds the size of the estimates given by the analysis of variance of Table 1.

To overcome this difficulty Patterson and Thompson (1971) proposed estimation procedures maximising the likelihood, not of the complete vector Y, but of all contrasts with zero expectation. We will refer to this procedure as modified maximum likelihood (MML). Patterson and Thompson (1971) were primarily concerned with the special case of incomplete block designs with possibly unequal block sizes. We now show that the method is applicable to the general model described in section 1 and that it is closely related both to MINQUE and to Nelder's method.

4. LIKELIHOOD OF ERROR CONTRASTS

First we derive an expression for the joint likelihood of all contrasts with zero expectation under the model defined by (1.1) and (1.2) and n-normal Y. Each such contrast is given by q^TY, where q is a vector in the column space of the matrix S given by

$$S = I - X(X^TX)^-X^T. \tag{4.1}$$

In equation (4.1) $(X^TX)^-$ can be any generalised inverse of X^TX but S itself is unique. It should be noted that $SX = 0$ so that $E(SY) = 0$. Also S is the projector onto the null space of X^T.

Every contrast q^TY can be expressed as the sum of two components, q_1^TY and q_2^TY, where q_1 is in the column space of SVS, and q_2 is in the null space. The component q_2^TY has zero variance, because $q_2^TVq_2 = q_2^TSVSq_2 = 0$, and hence is of no further interest in constructing the likelihood function. We call the other component q_1^TY an *error contrast*. It has zero expectation but non-zero variance.

Thus the required log likelihood of all q^TY, $q \in M(S)$, is also the log likelihood of all q_1^TY, $q_1 \in M(SVS)$; when Y is n-normal it is given by L, where

$$L = \text{constant} - \frac{1}{2}\ln|PVT^T| - \frac{1}{2}Y^TBY, \tag{4.2}$$

$$B = P^T(PVP^T)^{-1}P, \tag{4.3}$$

and P^T is any matrix of full column rank with the same column space as SVS.

MML estimates of θ_i are obtained by equating $\partial L/\partial\theta_i$ to zero. We note that P is independent of θ_i because (a) the design matrix X and hence also S and q are independent of θ_i, and (b) q_2 is independent of θ_i; otherwise the relationship $q_2SVS = 0$ would impose a constraint on the θ_i.

5. MML ESTIMATION

The MML equations of estimation are as follows :

$$Y^T BV_i BY = \text{trace } (BV_i),\ i = 1, 2 \cdots m, \tag{5.1}$$

with B as defined in (4.3). At first sight the solution may appear to depend through B on the choice of P. In reality B is the Moore-Penrose inverse of SVS and hence is unique i.e.

$$B = (SVS)^+. \tag{5.2}$$

This result can be obtained using the properties of the orthogonal projector S defined by (4.1) and another orthogonal projector S_* given by

$$S_* = P^T(PP^T)^{-1}P. \tag{5.3}$$

The operator S_* projects onto the column space of P^T. But this is also the column space of SVS so that alternative expressions for S_* are given by

$$S_* = SVS(SVS)^+ = (SVS)^+SVS. \tag{5.4}$$

It follows immediately that $S_*VS_* = SVS$. We need only show, therefore, that B is the Moore-Penrose inverse of S_*VS_*. The four conditions to be met are as follows :

(i) $S_*VS_*BS_*VS_* = S_*VS_*$, (5.5)

(ii) $BS_*VS_*B = B$, (5.6)

(iii) BS_*VS_* is symmetric,

(iv) S_*VS_*B is symmetric.

Now B and VB are unaltered by projection onto the column space of SVS so that

$$S_*B = B,\ S_*VB = VB, \tag{5.7}$$

Hence

$$BS_*VS_* = S_*VS_*B = S_*\,; \tag{5.8}$$

as S_* is symmetric conditions (iii) and (iv) are met. That conditions (i) and (ii) are also met can be verified by simple substitution for BS_*VS_* in the *LHS* of equations (5.5) and (5.6).

We further observe that the expectation of the term on the *LHS* of (5.1) is trace $(BVBV_i)$; reference to (4.3) shows that this simplifies to trace (BV_i) i.e. the term on the *RHS* of (5.1).

Hence MML estimation consists of equating the sums of squares $Y^T(SVS)^+V_i(SVS)^+Y$ to expectation, $i = 1, 2 \cdots m$.

6. SOLUTION BY METHOD OF SCORING

In the present paper we are not primarily concerned with practical procedures for solving equations (5.1). But one procedure — a straightforward application of Fisher's method of scoring — will be outlined as it reveals a relationship between MINQUE and MML. This procedure is very similar to that described by Anderson (1973) for maximising the complete likelihood.

We start with preliminary estimates $\tilde{\theta}_i$. These are substituted for θ_i in the formulae for V and B to give $\tilde{V}$ and $\tilde{B}$, where

$$\tilde{V} = V_1\tilde{\theta}_1 + V_2\tilde{\theta}_2 + \cdots + V_m\tilde{\theta}_m, \tag{6.1}$$

$$\tilde{B} = (S\tilde{V}S)^+. \tag{6.2}$$

Next we calculate the matrix $\{f_{ij}\}$ where

$$f_{ij} = \text{trace } (\tilde{B}V_i\tilde{B}V_j). \tag{6.3}$$

Improved estimates are given by $\tilde{\theta}'_i$ where

$$\tilde{\theta}'_i = \sum_j f^{ij}\boldsymbol{Y}^T\tilde{B}V\tilde{B}\boldsymbol{Y} \tag{6.4}$$

and

$$\{f^{ij}\} = \{f_{ij}\}^{-1}. \tag{6.5}$$

We note that

$$E(Y^T\tilde{B}V_i\tilde{B}Y) = \sum_j f_{ij}\theta_j. \tag{6.6}$$

Hence a single iteration of the MML scoring procedure consists of equating $Y^T(S\tilde{V}S)^+V_i(S\tilde{V}S)^+Y$ to expectation.

7. ESTIMATION OF β AND RESIDUALS

For the purposes of the present paper we suppose that whatever method is used for estimating the variance parameters, MINQUE, ML, MML or other, the elements of β are estimated as if V were known (apart possibly from a scaling factor). The appropriate method is given by Rao and Mitra (1971).

The estimates of β are given by the solution of

$$X^TV_*^+Y = X^TV_*^+X\hat{\beta}, \tag{7.1}$$

where

$$V_* = V \times kXX^T \tag{7.2}$$

and k is a scalar chosen so that $M(X) \subset M(V_*)$. In practice V is replaced by its estimate. If $M(X) \subset M(V)$ we take $k = 0$. This would be the case if, for example, V were positive definite.

We can now derive expressions for the residuals $Y - X\hat{\beta}$. These expressions will be useful in comparing MML with other estimation procedures. First we note that when $q^TX = 0$ and $q^TVS = 0$ the scalar $q^T\mathbf{Y}$ has zero expectation and zero variance i.e. $q^T\mathbf{Y} = 0$. Hence vectors b, c exist such that

$$\mathbf{Y} = Xb + VSc. \tag{7.3}$$

Premultiplication of both sides of (7.3) gives

$$X^TV_*^+Y = X^TV_*^+Xb \tag{7.4}$$

since

$$X^TV_*^+VS = X^TV_*^+V_*S = X^TS = 0. \tag{7.5}$$

Thus b is simply the estimate $\hat{\beta}$ given by the solution of (7.1) and the residuals are given by

$$R\mathbf{Y} = \mathbf{Y} - Xb = VSc, \tag{7.6}$$

where

$$R = I - X(X^TV_*^+X)^+X^TV_*^+. \tag{7.7}$$

The residuals can also be expressed in terms of $S\mathbf{Y}$. For premultiplication of both sides of (7.3) by S gives

$$S\mathbf{Y} = SVSc \tag{7.8}$$

so that

$$c = (SVS)^+\mathbf{Y}, \quad VSc = V(SVS)^+\mathbf{Y}, \tag{7.9}$$

and hence

$$R\mathbf{Y} = V(SVS)^+\mathbf{Y}. \tag{7.10}$$

When V is positive definite there are no constraints on y and equation (7.10) can be rewritten

$$R = V(SVS)^+. \tag{7.11}$$

Hence $(SVS)^+$ is given by

$$(SVS)^+ = V^{-1}R \tag{7.12}$$

$$\text{i.e. } (SVS)^+ = V^{-1} - V^{-1}X(X^TV^{-1}X)^+X^TV^{-1}. \tag{7.13}$$

8. MINQUE AND MML

In this section we show that Rao's (1970, 1973) MINQUE procedure is equivalent to a single iteration of MML as described in section 6.

The MINQUE estimators of θ_i are given by (3.1), where

$$A_i = B_M(\sum_j \lambda_{ij} V_j) B_M, \tag{8.1}$$

$$B_M = U^{-1} - U^{-1}X(X^T U^{-1} X)^+ X^T U^{-1} \tag{8.2}$$

U is defined in (3.2) and the λ_{ij} are chosen to ensure unbiasedness i.e.

$$\{\lambda_{ij}\} = \{\text{trace } (B_M V_i B_M V_j\}^{-1}. \tag{8.3}$$

Apart from substitution of U for V expression (8.2) for B_M is exactly the same as expression (7.13) for $(SVS)^+$ i.e.

$$B_M = (SUS)^+. \tag{8.4}$$

Thus MINQUE estimators are obtained by equating to expectation the sums of squares $Y^T(SUS)^+ V_i (SUS)^+ Y$, $i = 1, 2 \cdots m$. Comparison with section 6 shows that the MINQUE procedure is equivalent to one iteration of the MML scoring method with $\tilde{\theta}_i$ chosen in proportion to α_i^2. There are of course differences between the two methods. The final MML estimates are neither quadratic nor unbiased. MML estimation depends on the assumption of normality, MINQUE on choice of norm.

Hocking and Kutner (1973) have noted relationships between MINQUE, MML and a method proposed by Lamotte (1970).

9. MAXIMUM LIKELIHOOD

Similarly we can compare MML with ML, maximisation of the complete likelihood. Hartley and Rao (1967) have described the ML estimation procedure for Y n-normal, X of full column rank and V positive definite. The equations of estimation are

$$Y^T R^T V^{-1} V_i V^{-1} R Y = \text{trace } (V^{-1} V_i), \tag{9.1}$$

where

$$R = I - X(X^T V^{-1} X)^{-1} X^T V^{-1}. \tag{9.2}$$

Now $V^{-1}R$ is the Moore-Penrose inverse of SVS (equation (5.12)) so that (9.1) can be rewritten

$$Y^T (SVS)^+ V_i (SVS)^+ Y = \text{trace } (V^{-1} V_i). \tag{9.3}$$

Hence ML estimation consists of equating the sums of squares $Y^T(SVS)^+V_i(SVS)^+Y$ to trace $(V^{-1}V_i)$. In contrast MML equates the same sums of squares to their expectations, trace $\{(SVS)^+V_i\}$. We note that these expectations, unlike the quantities trace $(V^{-1}V_i)$ used in ML depend on the design matrix X.

10. ORTHOGONAL DESIGNS AND GENERALLY BALANCED DESIGNS

Finally, we show that MML estimation reduces to the classical analysis-of-variance method in orthogonal designs and to Nelder's (1968) method in designs with orthogonal block structure.

For both types of design the dispersion matrix takes the special form,

$$V = C_1\theta_1 + C_2\theta_2 + \cdots + C_m\theta_m. \tag{10.1}$$

where C_1, C_2 $\cdots$ C_m are mutually orthogonal indempotent matrices. These matrices define m strata of the analysis; in stratum i the projection C_iY is analysed.

Nelder (1965a) has shown that when $\Sigma C_i = I$ (so that V is p.d.) the null analysis of variance ignoring treatment effects consists of partitioning the sum of squares Y^TY as follows:

$$Y^TY = Y^TC_1Y + Y^TC_2Y + \cdots + Y^TC_mY. \tag{10.2}$$

A slight modification allows us to remove the limitation that V is p.d. We define $D = I - \Sigma C_i$ and replace Y^TY on the LHS of (10.2) by $Y^T(I - D)Y$. The term Y^TDY is constant because $DV = 0$ i.e. DY has zero variance. When $M(X) \subset M(V)$ the term Y^TDY is zero and (10.2) stands.

In an orthogonal design X is such that S commutes with each C_i. Hence the matrices SC_iS are also mutually orthogonal and idempotent and the residual sum of squares $Y^TS(I - D)SY$ can be partitioned as follows:

$$Y^TS(I - D)SY = Y^TSC_1SY + Y^TSC_2SY + \cdots + Y^TSC_mSY. \tag{10.3}$$

The analysis of variance estimates θ_i by the residual mean square in stratum i i.e.

$$\hat{\theta}_i = Y^TSC_iSY/\text{trace}(SC_iS). \tag{10.4}$$

We now show that these are MML estimates. The general MML method consists of equating $Y^T(SVS)^+C_i(SVS)^+Y$ to expectation. But for orthogonal designs SVS is given by

$$SVS = SC_1S\theta_1 + SC_2S\theta_2 + \cdots + SC_mS\theta_m. \tag{10.5}$$

As the SC_iS are mutually orthogonal and idempotent, the Moore-Penrose inverse of SVS is given by

$$(SVS)^+ = SC_1S\theta_1^{-1} + SC_2S\theta_2^{-1} + \cdots + SC_mS\theta_m^{-1}. \tag{10.6}$$

Hence

$$Y^T(SVS)^+C_i(SVS)^+Y = \theta_i^{-2}Y^TSC_iSY \tag{10.7}$$

so that MML, like the analysis of variance method, consists of equating Y^TSC_iSY to expectation.

By contrast the full maximum likelihood method (ML) equates Y^TSC_iSY to θ_i trace (C_i). As trace $(C_i) \geqslant$ trace (SC_iS) i.e. (the total number of d.f. in stratum i) $\geqslant$ (the number of error d.f. in stratum i) ML will in general underestimate θ_i. In the extreme case with $SC_i = 0$ there are no error degrees of freedom in stratum i. The analysis of variance and MML then give no estimate of θ_i whereas ML gives an apparent estimate of zero.

Nelder (1965b) has defined generally balanced designs as designs with X such that

$$X^TC_iX = \sum_k \lambda_{ik}Q_k, \quad i = 1, 2, \cdots m, \tag{10.8}$$

where the Q_k are mutually orthogonal idempotent matrices and the λ_{ik} are scalars. The dispersion matrix is again given by (10.1). For these designs Nelder (1968) proposed that θ_i should be estimated by equating the sums of squares $Y^TR^TC_iRY$ to expectation, with R as defined in (7.7).

We now show that MML gives exactly the same results. MML equates $Y^T(SVS)^+C_i(SVS)^+Y$ to expectation. As $VC_iV = C_i\theta_i^2$ this sum of squares can be rewritten $\theta_i^{-2}Y^T(SVS)^+VC_iV(SVS)^+Y$. Substituting RY for $V(SVS)^+Y$ (equation 7.10) we obtain $\theta_i^{-2}Y^TR^TC_iRY$ as yet another expression for the sum of squares. Hence MML also equates $Y^TR^TC_iRY$ to expectation.

It should be noted that condition (10.8) for generally balanced designs is not used in the above argument. Hence the principle of Nelder's (1968) method is appropriate for any design, balanced or otherwise, with spersion matrix given by (10.1). The generally balanced case is of course ery much simpler, both in terms of computation and interpretation.

REFERENCES

1. Anderson, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1**, 135—141.
2. Hartley, H. O. and Rao, J.N.K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika*, **54**, 93—108.
3. Hocking R. R. and Kutner, M. H. (1973). *Maximum likelihood analysis of mixed linear models: the balanced incomplete block model.* (Unpublished).
4. Lamotte, L. R. (1970). *A class of estimators of variance components.* University of Kentucky T.R. 10.
5. Mitra, S. K. (1971). Another look at the MINQUE of variance and covariance components. *Technical Report Math. Stat.* **3**, **71**, Indian Statistical Institute.
6. Nelder, J. A. (1965a). The analysis of randomized experiments with orthogonal block structure: Block structure and the null analysis of variance. *Proc. Roy. Soc. A*, **283**, 147—162.

7. NELDER, J. A. (1965b). The analysis of randomized experiments with orthogonal block structure : Treatment structure and the general analysis of variance. *Proc. Roy. Soc. A*, **283**, 163—178.
8. NELDER, J. A. (1968). The combination of information in generally balanced designs. *J.R. Statist. Soc. B*, **30**, 303—311.
9. PATTERSON, H. D. and THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545—554.
10. RAO, C. R. (1970). Estimation of heteroscedastic variances in a linear model. *J. Am. Statist. Assoc.* **65**, 165—172.
11. RAO, C. R. (1973). *Linear statistical inference and its applications* (Second Edition) Wiley, New York.
12. RAO, C. R. and MITRA, S. K. (1971). *Generalised inverse of matrices and its applications*. Wiley, New York.

4

Maximum likelihood estimation of variance components

by

R. Thompson

Math. Operationsforsch. Statist., Ser. Statistics, Vol. 11 (1980) No. 4, 545–561

Maximum Likelihood Estimation of Variance Components

ROBIN THOMPSON[1]

1. Introduction

This paper is concerned with the problem of estimating variance parameters in the mixed model by maximum likelihood (ML). The model can be specified by the following equations

$$\mathsf{E}(y) = X\alpha\,, \tag{1.1}$$

$$\mathrm{D}(y) = V = V_1\theta_1 + V_2\theta_2 + \ldots + V_m\theta_m\,. \tag{1.2}$$

The first equation gives the expected value of a vector y of observations in terms of a known $n \times t$ design matrix X (for simplicity we assume X is of full rank) and a vector of unknown parameters α. The second equation gives the dispersion matrix of y. Each element of the dispersion matrix is a linear function of m unknown variance parameters $\theta_1, \ldots, \theta_m$. The matrices V, $V_1, \ldots, V_m$ are symmetric and we assume that V is positive definite. We will discuss the case when

$$V_i = Z_i Z_i' \tag{1.3}$$

where Z_i is a $n \times q_i$ design matrix representing the allocation of q_i levels of a factor to the n observations and $Z_m = I$. The model for y can now be written as

$$y = X\alpha + Z_1 b_1 + \ldots + Z_m b_m \tag{1.4}$$

or

$$y = X\alpha + Z_1 b_1 + \ldots + e \tag{1.5}$$

where b_i is a vector of length q_i and $\mathsf{var}\,(b_i) = I\sigma_i^2$ and $\mathsf{cov}\,(b_i, b_j) = 0$ $(q_m = n\sigma_m^2 = \sigma^2)$. The parameters α are called fixed effects, the parameters b_i are called random effects and a model containing both fixed and random effects is called a mixed model. The parameters σ_i^2 represent variation in the random effects and are called variance components.

Interest in such a model and the estimation of variance components might arise for three different reasons.

---

[1] Paper given at the 4th International Summer School on Problems of Model Choice and Parameter Estimation in Regression Analysis, Mühlhausen (GDR), May 1979.

[2] Arc unit of statistics, university of Edinburgh, king's buildings, Edinburgh eh 9 3 jz, Scotland.

A. One might be interested in identifying the magnitude of the sources of variation in experiments or surveys. Indeed one of the first uses of variance components was in surveys (YATES and ZACOPANAY, 1935). (But the newer developments in variance component estimation have not been generated by or taken up in survey applications).

B. Interest in the efficient estimation of the fixed effects.

C. Interest in the prediction of the random effects.

Three examples which have partially generated my interest in variance component estimation illustrate these three different aims.

(1) *Recovery of inter block information in incomplete block experiments.* Often in designed experiments it is impossible to include all treatments in each block and so incomplete block designs are used. The main aim is B the efficient estimation of the fixed effects. Information is available on treatments from (i) contrasts within blocks, (ii) contrasts between blocks. In order to combine this information efficiently we need to know the relative sizes of the within and between block components.

(2) *Sire evaluation.* In several countries a large number of young dairy artificial insemination (AI) bulls are used for a limited time. Later, on the basis of their daughters' records a small proportion are used extensively. One model assumed is that a daughter yield has a herd-year-season effect and a random sire effect. Interest is then in predicting the yield of future daughters of the bulls i.e. the random sire effects (aim C). There is also interest in estimating the bull and residual variance components (aim A) so that the effects of different selection strategies might be investigated and optimal schemes implemented.

(3) *Variety trials.* In Great Britain cereal variety trials are carried out in order to give recommendations to farmers. Varieties are grown for a number of years on a number of sites. Not all varieties are grown in all years or at all sites. There is interest in estimating the fixed variety effects (aim B) and also in estimating the year, centre, year by centre, variety by year and variety by centre components (aim A). Estimates of these components are needed to make best use of the fairly scarce seed.

We will consider ML estimation of variance components and we will show that it gives a convenient framework for both the estimation of fixed effects and the prediction of random effects. HARVILLE (1977) has given a thorough review of this area and I have not very much to add to his comments. HARVILLE emphasises the general model (1.4) and most of the results are given in terms of this general model. This general formulation is heavily dependent on matrix formulae and so we use a succession of more limited model in an attempt to see more of the underlying structure.

2. Orthogonal block designs

Following PATTERSON and THOMPSON (1975) we first consider designs which the analysis of variance can give insight into estimators. We consider designs with orthogonal block structure (NELDER, 1965). These are designed such that the variance matrix is assumed to be, by randomization or other arguments, to be of the form

$$v=\sum_{i=0}^{k} C_i\theta_i \tag{2.1}$$

where C_i are idempotent, orthogonal matrices summing to the identity matrix. Hence the total sum of squares, $y'y$, can be separated into $k+1$ components $y'C_iy$ and we call $y'C_iy$ the sum of squares in the i-th stratum and the rank of C_i is the degrees of freedom associated with $y'C_iy$.

As an example consider a split-plot design with 3 blocks, 4 whole plots per block and 2 subplots per block then the variance matrix is

$$V=C_0\theta_0+C_1\theta_1+C_2\theta_2+C_3\theta_3 \tag{2.2}$$

where

$$\begin{aligned}
C_3&=I-Z_2(Z_2'Z_2)^{-1}Z_2',\\
C_2&=Z_2(Z_2'Z_2)^{-1}Z_2'-Z_1(Z_1'Z_1)^{-1}Z_1',\\
C_1&=Z_1(Z_1'Z_1)^{-1}Z_1'-J(J'J)^{-1}J,\\
C_0&=J(J'J)^{-1}J',
\end{aligned}$$

and Z_2 is a 24×12 design matrix with elements 0 and 1 representing the allocation of the 24 subplots to the 12 whole plots. Similarly Z_1 is a 24×3 matrix representing the allocation of subplots to the 3 blocks and J is a 24×1 matrix of ones. The sums of squares and degrees of freedom associated with each stratum are given in Table 1.

We can then try and get treatment estimates from stratum i using the transformed values C_iY. We will consider designs in which there is an orthogonal decomposition of the treatment effects and we can write α as

$$\alpha=\sum_{j=1}^{l} T_j\alpha \tag{2.3}$$

where T_j, like the C_i, are a set of orthogonal, idempotent matrices summing to the identity matrix. The information matrix for stratum i is $X'C_iX$ and the efficient estimation of treatment effects depends on finding a generalized inverse of $X'C_iX$. We will further restrict ourselves to designs, called generally balanced designs by NELDER (1965) such that

$$X'C_iX=\sum_{j=1}^{l}\lambda_{ij}T_j\,. \tag{2.4}$$

Hence a generalized inverse of $X'C_iX = \sum_{j=1}^{l} \lambda_{ij}^{-1} T_j$ where we interpret $\lambda_{ij}^{-1} = 0$ if $\lambda_{ji} = 0$. Then, if λ_{ij} is non-zero, we can estimate the treatment effects $T_j\alpha$ in the i-th stratum by

$$T_j\alpha_i = \lambda_{ij}^{-1} T_j X' C_i y \tag{2.5}$$

and the treatment effects $T_j\alpha$ sum of squares is $\lambda_{ij}(T_j\alpha_i)'(T_j\alpha_i)$ with rank (T_j) degrees of freedom. Whilst if $\lambda_{ij} = 0$ there is no information in the i-th stratum on the effect $T_j\alpha$. We note that the treatment effects are simple functions of $C_i y$, X' forms totals of observations and T_j takes deviations from means and λ_{ij}^{-1} is a scaling factor. If for some treatment effect, say $T_j\alpha$, there is only one non-zero λ_{ij}, say $\lambda_{i'j}$, then treatment effects $T_j\alpha$ can be thought of as being in stratum i'. When we can estimate each treatment effect $T_j\alpha$ $(j = 1, \ldots, l)$ in only one stratum (not necessarily the same stratum for all effects) then we have an orthogonal design and an analysis of variance can be written down subdividing each stratum sum of squares into treatment sum of squares and error sum of squares. Each treatment factor sum of squares occurs in only one stratum. The variance parameter θ_i can be estimated by the stratum error mean square (if there is one).

For example in the split-plot example suppose there are four levels on one treatment factor (A) allocated to whole plots and two levels of a second factor (B) allocated to subplots. Then the analysis of variance in Table 1 can be constructed. For completeness we include the correction factor for the mean $y'C_0y$. The parameter θ_1 is then estimated by the blocks mean square, θ_2 by the whole plot error mean square and θ_3 by the subplot error mean square.

Table 1

Analysis of variance of results of split-plot experiment

	d.f.	d.f.	S.S.	M.S. estimate
Mean stratum	1		$y'C_0y$	
Block stratum	2		$y'C_1y$	θ_1
Whole plot stratum	9		$y'C_2y$	
Factor A		3	S_A	
Error		6	E_W	θ_2
Subplot stratum	12		$y'C_3y$	
Factor B		1	S_B	
A × B		3	S_{AB}	
Error		8	E_S	θ_3

This method is well known. However there can be difficulties when there are departures from orthogonality of block structures and treatment structures. Analysis of variance methods can still be used in which mean squares are equated to expectation. Indeed the very popular methods of HENDERSON (1953) are

based on this principle. However in general these methods are neither unique or efficient.

We now consider the case when there are estimates of treatment effects $T_j\alpha$ in at most two strata (not necessarily the same two strata for all effects). Suppose there is information on effect $T_j\alpha$ in strata r and s then the difference between the estimates is $T_j(\alpha_r-\alpha_s)$ and the sum of squares of these differences $(\alpha_r-\alpha_s)'\times$ $\times T_j'T_j(\alpha_r-\alpha_s)$ has degrees of freedom rank (T_j) and expectation $\lambda_{rj}^{-1}\theta_r+\lambda_{sj}^{-1}\theta_s$. Other information on θ_r is available from the error sum of squares from stratum r. The problem is to combine this information together to form an efficient estimate for the variance parameters. If we assume that y has a multivariate normal distribution then the information on the θ's can be thought of as most $m+1+l$ observations from gamma distributions with expectation a linear combination of the variance parameters. This corresponds to a generalized linear model (NELDER and WEDDERBURN, 1972) and they show that ML estimates of the variance parameters can be thought of as being derived by an iterative weighted least squares.

We briefly give their method as we wish to discuss generalizations of the mixed models later. They consider models of the form

$$\mathsf{E}(y)=\mu$$

where $\mu_i=h(\eta_i)$, $\eta=X\beta$ and μ, y and η are vectors of length n with elements μ_i, y_i and η_i. X is an $n\times p$ matrix containing p independent variates and we wish to estimate the p parameters in the vector β. The mean value of y_i is a function, h, of η_i which is a linear function of the independent variables.

NELDER and WEDDERBURN assumed that the distribution of y belongs to the exponential family of densities of the form

$$p(y,\theta,\Phi)=\exp\left[\{y\theta-b(\theta)+d(y)\}/a(\Phi)+t(\Phi,y)\right]$$

where Φ is a scale parameter. Since μ, the mean of y, can be written as $b'(\theta)$ there is a relationship between μ and θ. For the Normal, Poisson and gamma distributions (the one we want to use), all of which are included in the family, we have $b(\theta)={}^1/_2\theta^2$, $\exp(\theta)$ and $\ln\theta$ and $\mu=\theta$, $\exp(\theta)$ and $-\theta^{-1}$ respectively.

NELDER and WEDDERBURN show that the solution of the ML equations for β is equivalent to an iterative weighted least squares procedure, in each iteration minimizing $(z-X\beta)'W(z-X\beta)$ with respect to $\beta\cdot z$ is a modified dependent variable with elements given by

$$z_i=\eta_i+(y_i-\mu_i)\cdot(\mathrm{d}\eta_i/\mathrm{d}\mu_i)$$

and W is a diagonal matrix with elements given by $w_i=(\mathrm{d}\mu_i/\mathrm{d}\eta_i)^2/\tau_i^2$ and μ, η, W are calculated affresh in each iteration using the current values of β and $\tau_i^2=\mathrm{d}\mu_i/\mathrm{d}\theta_i$.

In our case we take y_i to be m_i the i-th mean square providing information on the variance parameters of interest (θ_i corresponds to β_i in the above argument). Obviously if only one mean square provides information on variance θ_i then there is no need for the procedure for estimating θ_i. As $\mu_i=\eta_i$ then z_i reduces to m_i and $w_i=d_i/2\mu_i^2$ where d_i is the degrees of freedom associated with m_i.

The simplest example is perhaps balanced incomplete blocks. Let the design have t treatments each replicated r times in b blocks of k plots each. Every pair of treatments occur together in λ blocks. The treatment model can be written as $\alpha = T_0\alpha + T_1\alpha$ with T_1 representing differences about the mean. It can be shown that $\lambda_{11} = r(1-E)$, from the block stratum, and $\lambda_{12} = rE$, from the within block stratum, where E is the efficiency factor $= t(k-1)/[k(t-1)]$. Hence the analysis of variance given by subdivision (1) in Table 2 can be constructed. YATES (1940) gives this and an alternative subdivision ((2) in Table 2) based on fitting treatments and then blocks. The error sums of squares are the same in the two subdivisions. YATES suggests using the block eliminating treatment sum of squares to estimate the block component σ_b^2 (where $\sigma^2 = \theta_2$, $\sigma^2 + k\sigma_b^2 = \theta_1$). In our method we use the sum of squares of differences in treatment between strata and this is proportional to YATES's Blocks (eliminating treatments) treatment component. (The expectation of our mean square is $\theta_1/r(1-E) + \theta_2/rE = [E\theta_1 + (1-E)\theta_2]/[rE(1-E)] = [\sigma^2 + kE\sigma_b^2]/[rE(1-E)]$.) Hence we use the same sums of squares as YATES but weight them differently.

Table 2

Analysis of variance for balanced incomplete block designs

Subdivision (1)	d.f.	Subdivision (2)	Expectation of M.S.
Mean Stratum	1	Mean Stratum	
Block Stratum		Blocks (eliminating treatments)	
Treatment component	$t-1$	Treatment component	$\sigma^2 + kE\sigma_b^2$
Error	$b-t$	Remainder	$\sigma^2 + k\sigma_b^2$
Total	$b-1$	Total	$\sigma^2 + [(bk-t)/(b-1)]\,\sigma_b^2$
Within Block Stratum			
Treatments	$t-1$	Treatments (ignoring blocks)	
Error	$rt-t-b+1$	Intra-block error	σ^2

The procedure can be extended to the case when treatment estimates for effect $T_j\alpha$ are available on s_j+1 strata. We can construct s_j differences between the first s_j estimates and the last estimate and hence calculate M_j a $s_j \times s_j$ matrix representing mean squares and products of these differences. Again the expectation of M_j can be written as a linear function of the variance parameters, i.e. $\sum_{i=0}^{m} X_{ji}\theta_i = \mu_j$ where X_{ji} $(i = 0, \ldots, m)$ and μ_j are now $s_j \times s_j$ matrices.

The log-likelihood of matrices providing information on the variance parameters can be written as

$$\mathfrak{L}' = \text{const} - \tfrac{1}{2}\sum_{h=1}^{n} d_h\,[\ln|\mu_h| + \text{tr}\,(M_h\mu_h^{-1})]\,. \tag{2.6}$$

The values of θ_i that maximize (2.6) satisfy

$$\frac{d\mathfrak{L}'}{d\theta_i} = \sum_{h=1}^{n} d_h \operatorname{tr}(\mu_h^{-1} M_h^{-1} \mu_h^{-1} X_{hi}) - \sum_{h=1}^{n} d_h \operatorname{tr}(\mu_h^{-1} X_{hi}) = 0 . \tag{2.7}$$

Usually an iterative solution to (2.7) is needed. One based on using the expectations of the second differentials is very similar to weighted least squares is suggested by ANDERSON (1973). In this scheme θ_i is estimated from

$$\sum_{j=0}^{m} \tilde{A}_{ij} \hat{\theta}_j = \tilde{\beta}_i , \tag{2.8}$$

where

$$\tilde{A}_{ij} = {}^1/_2 \sum_{h=1}^{n} d_h \operatorname{tr}(\tilde{\mu}_h^{-1} X_{hi} \tilde{\mu}_h^{-1} X_{hj}) , \quad \tilde{B}_i = {}^1/_2 \sum_{h=1}^{n} d_h \operatorname{tr}(\tilde{\mu}_h^{-1} X_{hi} \tilde{\mu}_h^{-1} M_h) \tag{2.9}$$

and $\tilde{\mu}_h$ is an initial estimate of μ_h. The procedure can be repeated using $\hat{\theta}_i$ to give $\tilde{\mu}_h$ until the estimates converge. We note that using $\mu_h = \sum_{j=0}^{m} X_{hj}\theta_j$ quickly derives (2.9) from (2.8). Equation (2.9) can be thought of the NELDER and WEDDERBURN approach for gamma variables. In (2.9) $\tilde{\mu}_h^{-1}$ is used as a weighting matrix and ${}^1/_2 d_h \operatorname{tr}(\tilde{\mu}_h^{-1} X_{hi} \tilde{\mu}_h^{-1} X_{hj})$ replaces ${}^1/_2 d_h X_{hi} X_{hj} / \tilde{\mu}_h^2$ in the gamma variable case.

This development has been motivated by generally balanced designs but (2.9) can be useful in other circumstances. THOMPSON (1976a, b) has given examples in a genetic context. These arise when there is an interest in estimating variance parameters and a partition of variance into covariances between relatives is possible but there are more covariances than parameters of interest.

It should be noted that when estimating the variance parameters we have not been maximizing the log-likelihood of y, $\mathfrak{L}$, but the log-likelihood of error sums of squares and cross products. For example in the split-plot example using the least squares estimates of treatments we find

$$\mathfrak{L} = -{}^1/_2 [\ln \theta_0 + 2 \ln \theta_1 + 9 \ln \theta_2 + 12 \ln \theta_3 + y' C_1 y/\theta_1 + E_w/\theta_2 + E_s/\theta_3]$$

and

$$\mathfrak{L}' = -{}^1/_2 [2 \ln \theta_1 + 6 \ln \theta_2 + 8 \ln \theta_3 + y' C_1 y/\theta_1 + E_w/\theta_2 + E_s/\theta_3] .$$

The coefficient of $\ln \theta_i$ in $\mathfrak{L}$ is the stratum degrees of freedom and in $\mathfrak{L}'$ it is the stratum degrees of freedom minus the number of treatment effects estimated in that stratum. The terms dependent on the data are the same in both expressions. Hence it seems more natural to me to use $\mathfrak{L}'$ as it takes account of estimating treatment effects and will give the analysis of variance estimators for orthogonal designs. In this example $\mathfrak{L}'$ gives the analysis of variance estimators for θ_1, θ_2 and θ_3 whilst using $\mathfrak{L}$ gives estimates of θ_2 and θ_3 only two-thirds of the analysis of variance estimators. The case of θ_0 corresponds to the extreme case of one observation from a normal distribution. Using $\mathfrak{L}'$ suggests θ_0 is not estimable (essentially $C_0 y$ is used to estimate the mean value) whilst using $\mathfrak{L}$ suggests estimating θ_0 by zero.

3. Incomplete block designs with unequal sizes

In this section we use the model generated by these designs and set up the ML equations for estimating variance parameters and interpret these equations. In some circumstances for instance animal experiments where blocks consist of groups of similarly related animals it often happens that blocks are not of equal size and the variance structure implied by the orthogonal block structure is not appropriate. The orthogonal block structure generates

$$V = C_0\theta_0 + C_1\theta_1 + C_2\theta_2 \tag{3.1}$$
$$= I\theta_2 + Z_1 (Z_1'Z_1)^{-1} Z_1' (\theta_1 - \theta_2) + J (J'J)^{-1} J' (\theta_0 - \theta_1)$$

and it is more usual to assume

$$V = I\sigma^2 + ZZ'\sigma_b^2 = (I + ZZ'\gamma)\,\sigma^2, \tag{3.2}$$

or equivalently

$$y = X\alpha + Zb + e, \tag{3.3}$$

where Z is an $n \times q$ design matrix.

Equation (3.2) implies the covariance between observations in the same block are the same in all block independent of block size. Whilst (3.1) implies that the covariance between observations in the same block is inversely proportional to the block size. We will assume variance structure (3.2) in this section. Equation (3.1) might be more appropriate in some field experiments but I doubt if the difference between (3.1) and (3.2) would ever be important in these experiments.

In section 2 we intuitively set up sums of squares and cross products of residuals but in this case it is not immediately obvious how to do this. PATTERSON and THOMPSON (1971) suggest maximizing the likelihood of error contrasts i.e. contrasts with zero expectation with non-zero variance to estimate the variance components. This likelihood corresponds to the likelihood $\mathfrak{L}'$ used in section 2 and again we call this likelihood $\mathfrak{L}'$. This approach is called a restricted maximum likelihood approach by HARVILLE (1977) after W. A. THOMPSON (1962) who used $\mathfrak{L}'$ by 'restricting' himself to using the likelihood of the location invariant part of the data in balanced designs. If we consider Sy where

$$S = I - X (X'X)^{-1} X \tag{3.4}$$

then we see this represents error contrasts since $\mathsf{E}(Sy) = 0$. The elements of Sy are deviations from treatment means. The likelihood of the error contrasts cannot be written down immediately because S is of rank $n-t$. There are several ways of deriving $\mathfrak{L}'$ perhaps the nearest is to note that $\mathfrak{L} = \mathfrak{L}' + \mathfrak{L}''$ where $\mathfrak{L}''$ is the likelihood of $X'V^{-1}y$ (PATTERSON and THOMPSON, 1971). Hence $\mathfrak{L}' = \mathfrak{L} - \mathfrak{L}''$

$$\begin{aligned} &= -\tfrac{1}{2} \ln |V| - \tfrac{1}{2} (y - X\alpha)' V^{-1} (y - X\alpha) + \tfrac{1}{2} \ln |X'V^{-1}X| \\ &\quad + \tfrac{1}{2} (y - X\alpha)' V^{-1} X (X'V^{-1}X)^{-1} X'V^{-1} (y - X\alpha) \\ &= -\tfrac{1}{2} \ln |V| + \tfrac{1}{2} \ln |X'V^{-1}X| - \tfrac{1}{2} (y - X\hat{\alpha})' V^{-1} (y - X\hat{\alpha}), \end{aligned} \tag{3.5}$$

where $\hat{\alpha}$ the least squares estimate of α satisfies

$$(X'V^{-1}X)\,\alpha = X'V^{-1}y\,. \tag{3.6}$$

An alternative derivation in terms of a generalised inverse of (SVS) or the latent roots of SVS given by PATTERSON and THOMPSON (1971) is interesting in that it links with the method of the previous section. They show that $\mathfrak{L}'$ corresponds to the likelihood of $u = P'y$ where P an $n \times n-t$ matrix is equal to AB where A is chosen so that $S = AA'$ and $AA' = I$ and B is chosen so that $B'A'VAB = I$. Hence $PP' = S$, $P'P = I$ and $P'HP = \mathrm{diag}\,(\xi_s) = \sigma^2 + \lambda_s\sigma_b^2$. Hence $\mathfrak{L}'$ can be thought of as $(n-t)$ independent sum of squares $u_s'u_s$ with expectation $\sigma^2 + \lambda_s\sigma_b^2$ and the generalized linear model scheme of section 2 could be used. It is not a very useful form because it involves the calculation of P. However some matrix manipulations show that only $q-1$ of the λ_s are non-zero (equal to the latent roots of $Z'SZ$) and the sums of squares for the u_s with zero λ_s can be written $y'Sy - y'SZ(Z'SZ)^- Z'Sy$ or the residual sum of squares after fitting treatment and blocks. It is difficult if not impossible to generalize this argument to include another block structure with unequal block sizes (given by Z_1) as it is then impossible to find a P to simultaneously diagonalize I, ZZ' and Z_1Z_1'.

More useful computational forms for estimating the variance parameters can be derived from manipulating P or more directly by first differentiating (3.5). The differentials are functions of V^{-1} (for instance $\mathrm{d}\mathfrak{L}'/\mathrm{d}\sigma_b^2$ contains $(y - X\hat{\alpha})' \times V^{-1}ZZ'\,V^{-1}\,(y - X\alpha)$) and simplifications arise from using the matrix identity

$$H^{-1} = I - Z\,(Z'Z + \Gamma^{-1})^{-1}\,Z' \quad \text{if} \quad H = I + Z\Gamma Z'\,. \tag{3.7}$$

This enables $\hat{\alpha}$ the solution of (3.6) to be written as the solution of

$$\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + I\gamma^{-1} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} X'y \\ Z'y \end{pmatrix} \tag{3.8}$$

(HENDERSON in HENDERSON et al., 1959).

The differentials of $\mathfrak{L}'$ with respect to σ^2 and γ (we use this parametrization because it leads to more compact expressions) are (from PATTERSON and THOMPSON, 1971 and HARVILLE, 1977).

$$2\,\frac{\mathrm{d}\mathfrak{L}'}{\mathrm{d}\sigma^2} = \frac{-(n-t)}{\sigma^2} + \frac{R}{\sigma^4}\,, \qquad 2\,\frac{\mathrm{d}\mathfrak{L}'}{\mathrm{d}\gamma} = -\mathrm{tr}(U) + \frac{B}{\sigma^2}\,, \tag{3.9}$$

$$2\,\frac{\mathrm{d}^2\mathfrak{L}'}{\mathrm{d}\sigma^4} = \frac{n-t}{\sigma^4} - \frac{2R}{\sigma^4}\,, \qquad 2\mathsf{E}\left(\frac{\mathrm{d}^2\mathfrak{L}'}{\mathrm{d}\sigma^4}\right) = \frac{-(n-t)}{\sigma^4}\,,$$

$$2\,\frac{\mathrm{d}^2\mathfrak{L}'}{\mathrm{d}\sigma^2\mathrm{d}\gamma} = \frac{-B}{\sigma^4}\,, \qquad 2\mathsf{E}\left(\frac{\mathrm{d}^2\mathfrak{L}'}{\mathrm{d}\sigma^2\mathrm{d}\gamma}\right) = \frac{-\mathrm{tr}(U)}{\sigma^2}\,, \tag{3.10}$$

$$2\,\frac{\mathrm{d}^2\mathfrak{L}'}{\mathrm{d}\gamma^2} = \mathrm{tr}\,(U^2) - Z\beta' U\beta/\gamma^2\sigma^2, \qquad 2\mathsf{E}\left(\frac{\mathrm{d}^2\mathfrak{L}'}{\mathrm{d}\gamma^2}\right) = -\mathrm{tr}\,(U^2)$$

where

$$B = \beta'\beta/\gamma^2, \quad R = y'Sy - y'SZ\beta\,, \tag{3.11}$$

$$W = (Z'SZ + I\gamma^{-1})^{-1} \quad \text{and} \quad U = I\gamma^{-1} - \gamma^{-1}W\gamma^{-1}\,. \tag{3.12}$$

All the terms in (3.9) and (3.10) arise naturally from the solution of (3.8). The bottom right corner of the inverse of the matrix on the left-hand side of (3.8) is W. If $\gamma^{-1}=0$ in (3.8) then (3.8) reduces to least squares equations assuming fixed block effects and R corresponds to the residual sum of squares after fitting blocks and treatments. Estimates of σ^2 and γ that maximize $\mathfrak{L}'$ therefore satisfy

$$R=(n-t)\,\sigma^2 \quad \text{and} \quad \beta'\beta=q\sigma_b^2-\sigma^2\,\mathrm{tr}\,(W)\,. \tag{3.13}$$

These can be thought of as analysis of variance estimators in the sense that $\mathsf{E}(R)=(n-t)\,\sigma^2$ and $\mathsf{E}(\beta'\beta)=q\sigma_b^2-\sigma^2\,\mathrm{tr}\,(W)$ (when R and $\beta'\beta$ are expressed in terms of y, σ^2 and σ_b^2).

Not only does (3.8) give a convenient way of estimating α but as HENDERSON (1963) has pointed out β is not just a computational artefact but is useful when one wishes to predict functions of the random effects. The sire evaluation case in the introduction is one example. In this case a model of the form (3.2) or (3.3) is sometimes used then σ_b^2 is the covariance between daughters of the same bull (the half-sib covariance) and one is interested in predicting b. The problem of predicting random effects has received less attention than the problem of estimating fixed effects.

HENDERSON (1973) consider the problem of predicting a linear function of α and b say $w_i=\lambda_1'\alpha+\lambda_2'b$. He suggested using a predictor $\hat{w}_i$ for w_i where (i) $\hat{w}$ is a linear function of y (ii) $\hat{w}_i$ is an unbiased estimator of w_i and (iii) the mean square error $(\mathsf{E}\,(\hat{w}_i-w_i)^2)$ is minimized. HENDERSON showed that this best linear unbiased predictor (BLUP) (for known V) satisfied

$$\hat{w}_i=\lambda_1'\hat{\alpha}+\lambda_2'\sigma_b^2 Z'V^{-1}\,(y-X\hat{\alpha}) \tag{3.14}$$

and that this reduces to

$$\hat{w}_i=\lambda_1'\hat{\alpha}+\lambda_2'\beta \tag{3.15}$$

where $\hat{\alpha}$ and β satisfy (3.8) and so β can be thought of as a predictor of u. The $\sigma_b^2 Z'V^{-1}$ term can be thought of as regression coefficients as it represents the covariance of u and y divided by the variance of y. When α is known the BLUP procedure is equivalent to the use of selection indices (HAZEL, 1943) to combine the information on relatives. In the sire evaluation case both methods suggest predicting a bulls merit from n daughters by

$$n/(n+\gamma^{-1})\times\text{the bull's daughter average corrected yield.} \tag{3.16}$$

When α is unknown some sire evaluation methods (THOMPSON, 1979) act by first assuming the bull effects are fixed effects and then estimate the bull effects and then use these estimates in a formula similar to (3.16). Implicitly then the bull effects are first assumed fixed effects and then random effects. I much prefer the BLUP approach $\beta'\beta$ can then be seen to be quite simply a sum of squares of predicted values. ROBERTSON (1962) suggested using $\beta'\beta$ to estimate σ_b^2 in a one way classification but he interpreted it as a weighted sum of squares of block means.

The generalization to include more block structures is fairly straightforward (PATTERSON and THOMPSON, 1971, HARVILLE, 1977). Again the key equations can

be thought of as least squares equations with the addition of diagonal terms. The inverse of the matrix involved in these equations can be partitioned and these partitioned matrices contribute to the first and second differentials. Another generalization is to deal with q traits and then one is interested in estimating the $q \times q$ matrices of residual variances and covariances, Σ, and the block variances and covariances, Σ_b (THOMPSON, 1973). Then the key equations can be thought of as q sets of least squares equations for the q traits and extra terms for the block predictors that link the q sets together. The computation can be simplified by finding a transformation T such that

$$T\Sigma_b T' = I \quad \text{and} \quad T\Sigma T' = D \tag{3.17}$$

where D is a diagonal matrix. The equations for the canonical variates derived using T, then split into q parts each similar to (3.8). Unfortunately such a subdivision is not in general possible if there is more than one block structure.

4. Solution of likelihood equations

HARVILLE (1977) has discussed the iterative solution of the likelihood equations. I have found a scheme based on the method of scoring using the expected values of the second differentials satisfactory. For the model (3.2) first F_0, the information matrix of σ_b^2 and σ^2, is found from (3.10) and $2d_0' = (B/\sigma^4,\ R/\sigma^4 - B\sigma_b^2/\sigma^2)$ is calculated using initial estimates σ_0^2 and σ_{b0}^2 for σ^2 and σ_b^2. New estimates of σ^2 and σ_b^2 can be found from

$$F_0 \begin{pmatrix} \sigma_b^2 \\ \sigma^2 \end{pmatrix} = d_0 \tag{4.1}$$

and the procedure repeated until the estimates converge. The generalized linear model scheme in section 2 also uses the method of scoring. It can be shown that for any choice of σ^2 and σ_{b0}^2 (4.1) can be thought of as equating d_0 to its expected values (HARVILLE, 1977). If the terms in (3.9) can be formed then there is no real difficulty in forming either the second differentials or their expected values and so the Newton-Raphson scheme could be implemented just as easily as (4.1).

The estimator given by (4.1) has been derived using likelihood ideas but alternative arguments lead to (4.1). The locally best (minimum variance) translation invariant quadratic estimator at $\sigma^2 = \sigma_0^2$ $\sigma_b^2 = \sigma_{b0}^2$ is given by (4.1) when b and e are normally distributed (LA MOTTE, 1973). RAO's (1973) method of minimum norm quadratic unbiased estimation gives another justification without any normality assumption. RAO's estimators are quadratic and unbiased and so for model (3.2) are of the form $y'S_1y = \hat{\sigma}^2$ and $y'S_2y = \hat{\sigma}_b^2$ where $\mathsf{E}(y'S_1y) = \sigma^2$ and $\mathsf{E}(y'S_2y) = \sigma_b^2$ and S_1 and S_2 are symmetric matrices. RAO suggests choosing S_1 and S_2 so that the Euclidean Norm (sum of squares of all the elements) of S_1V^* and S_2V^* are minimized, where $V^* = \theta_1^* I + \theta_2^* ZZ'$. The MINQUE estimators of σ^2 and σ_b^2 satisfy (4.1) if σ_0^2 and σ_{b0}^2 are replaced by θ_1^* and θ_2^*. The estimators depend on θ_1^* and

θ_2^* and RAO (1973) recommends that, when possible, they should be chosen proportional to σ^2 and σ_b^2. Whilst the MINQUE estimators have some intuitive appeal and reduce to other estimators when normality is assumed (and have generated a lot of interest) the use of the measure is not accepted or understood by all statisticians.

Another scheme is based on manipulating (3.13) into the form $\sigma_b^2 = (\beta'\beta + \sigma^2 \operatorname{tr} W)/q$ or $\sigma_b^2 = \beta'\beta/(q - \operatorname{tr}(W)/\gamma)$ (HARVILLE's (6.3) and (6.5)) and whilst appealing (similar schemes are suggested by NELDER (1968) and THOMPSON (1969)) I have found that they can be slow to converge. This scheme is mentioned by DEMPSTER, RUBIN and LAIRD (1977) as a special case of the EM algorithm. This 'algorithm' is a method of computing maximum likelihood estimates from incomplete data. They treat the effects of the levels of the random factors as missing values. They give formulae which enable the rate of convergence to be calculated. For the one-way classification with n observations per block then the rate of convergence is approximately $\max(1 - \{n\gamma/(n\gamma+1)\}^2, 1/n)$. This tends to 1 as $n\gamma$ tends to zero. On the other hand if $n\gamma$ is large the rate of convergence is of order $1/n$ and convergence should be rapid.

One problem with these schemes is that they need the calculation of W the inverse of $Z'SZ + I\gamma^{-1}$. If there are a large number of random effects this might be difficult. Sometimes results about partitioned matrices reduce the problem to a manageable size. For instance in the variety trial example partitioning the effects into those associated with varieties and those not make it feasible to estimate the variance components (THOMPSON, 1977).

I have little to say about constraints primarily because I tend to think of the variance models as (1.2) so that the only constraint needed would make V positive definite. Using a mixed model approach one can argue that the variance components should be non-negative. HARVILLE (1977) has considered how solutions might be found that satisfy these constraints.

The multivariate version of the constraint $\sigma_b^2 > 0$ is that Σ_b is positive definite, i.e. any linear combination of the q traits has a positive block component. HILL and THOMPSON (1978) have recently worked out the probability that analysis of variance estimates of Σ_b are not positive definite for some one-way classifications. The probabilities can be surprisingly high, for example with 160 groups of size 10 then this probability approaches one if there are 8 independent traits with intraclass correlation 0.025 or if there are 14 independent traits with intraclass correlation 0.0625.

5. Transformations

HARVILLE (1977) pointed out that the performance of the numerical algorithms might be improved by making the likelihood function more quadratic. HEMMERLE and HARTLEY (1973) suggest making a transformation that reduces to using the

parameters $\sqrt{\gamma}$ and σ^2 for the model (3.3). The transformation was introduced so that $\hat{\sigma}_b^2 \geqq 0$ and they suggest that this transformation improves the convergence of the iterative procedure.

Another class of transformations is suggested by consideration of orthogonal designs. Then we see each stratum variance is estimated independently of each other and the information matrix for the stratum variances is diagonal with elements half the degrees of freedom divided by the stratum variance squared. If θ represents the vector of variance components (in the same order as in the analysis of variance) then $T\theta$ represents the stratum variances and T is an upper triangular matrix with elements in the last column 1 (the coefficient of the residual variance).

This suggests that if one has calculated F the information matrix for variance components in an unbalanced design then approximate stratum variances could be found from $T\theta$ where

$$TFT' = D \tag{5.1}$$

where D is a diagonal matrix and again T is an upper triangular matrix with elements in the last column 1. In the orthogonal case the stratum variances have a gamma distribution so obvious transformations are log $(T\theta)$ (to make the variances of the stratum variances independent of their mean value) $(T\theta)^{3/7}$ (to improve the quadratic approximation) or even $(T\theta)^{1/3}$ (the power used in the Wilson and Hilferty (1931) transformation). I have not investigated any of these possibilities, apart from plotting several likelihood surfaces for several examples and verifying that taking log and square root transformations make the surface more symmetrical about the maximum likelihood estimates.

6. Other models

The emphasis in this paper has been on variance components in linear models assuming some of the effects are normally distributed. Of course there are other situations when one wishes to estimate variance parameters. One I have been interested in is when there is an underlying variate, η, normally distributed with mean $X\alpha$ and variance $I+ZZ'\sigma_b^2$ and the measured variate, y, is one if $\eta \geqq 0$ and zero if $\eta < 0$. Curnow and Smith (1975) have discussed the application of such a model in genetic situations. ML estimation involves the evaluation of probabilities from truncated multivariate distributions. The work of Dempster et al. (1977) suggests thinking of the model as $y = \Phi(X\alpha + Zb)$, where Φ is the normal integral, y is binomially distributed and b has variance $I\sigma_b^2$. If b were fixed effects then this is a generalized linear model and the method of section 2 could be used. This suggests modifying the weighted least squares equations by the addition of diagonal terms, just as (3.8) can be thought of as modified least squares equations. If this procedure works then presumably it could be

applied to any generalized linear model extending them from fixed effects models to mixed effects models. One could perhaps generalize the method even more and allow a transformation $h_1(b)$ to make b Normally distributed.

LEONARD (1972) has suggested a similar method for binomial data assuming a logistic model. However this method in several numerical examples gives estimates of zero for σ_b^2 in cases that I think a non-zero estimate would be more appropriate.

To explain this we need to go back to the paper of LINDLEY and SMITH (1972), who consider a simpler model than LEONARD (1972) but from a similar Bayesian viewpoint. One model considered by LINDLEY and SMITH is of the form $y = X\alpha + Zb + e$ with $m = 2$ and e having a normal distribution with variance σ^2. LINDLEY and SMITH argue that in some circumstances we might have prior information that the α and b parameters are exchangeable (perhaps in subsets). This implies a priori that the elements of α and b are normally distributed with variances σ_α^2 and σ_b^2. They also argue that there is sometimes prior information on the variance parameters σ_α^2, σ_b^2 and σ^2 and use χ^2 distributions with v_α, v_b and v degrees of freedom to characterize this information. For known values of σ_α^2, σ_b^2 and σ^2 LINDLEY and SMITH show that the posterior mode α, b can be found from equations similar to (3.8). We note that this gives a natural interpretation to (3.8) as the combination of information from the data and the prior information. To estimate the variance parameters LINDLEY and SMITH suggest, as an approximation, using the mode of the posterior distribution of (α, b, σ_α^2, σ_b^2 and σ^2). This leads to equation of the form ($\beta'\beta$ + prior information) $= (q + v_b)\,\sigma_b^2$. This is similar to (3.13) but is simpler in that no account is taken of predicting β.

Let us consider the one-way classification model with q groups of n. If we write $s = n\sigma_b^2/(\sigma^2 + n\sigma_b^2)$, the shrinkage factor used in (3.16), then it can be shown that the estimate of s satisfies a cubic equation in terms of the sums of squares within and between groups, n, q and the prior information.

A less committed Bayesian is probably more interested in the limiting case $v \to 0$ $v_b \to 0$ (vague information on the variance parameters), $\sigma_\alpha^2 \to \infty$ (vague prior information on the fixed effects) which corresponds to the usual mixed model. LINDLEY and SMITH argue that $v_b = 0$ should not be used but they and several of their co-workers have used $v_b = 0$ in numerical examples (for instance NOVICK, JACKSON and THAYER (1971), SMITH (1973)). This special case is interesting. The one root is always zero corresponding to a shrinkage factor of zero, as noted by HARVILLE (1977). There can be two positive roots, presumably the bigger corresponds to a mode, but this depends on the sizes of the sums of squares. Let F be the ratio of the between group mean square to the within groupmean square. Then there is a positive root if $F > 4q^*\,(q^*n^* + q^*)/q^{*2}\,n^{*2} = c$ where $q^* = q + 2$, $n^* = n + 2$. In the numerical examples cited above presumably this condition (or similar conditions) was satisfied.

We might compare this approach with a significance test that $\sigma_b^2 = 0$. Then the hypothesis $\sigma_b^2 = 0$ is accepted, leading to a shrinkage factor of 0, if F is less than

a critical value depending on the level of significance and the degrees of freedom involved. We note that the significance test critical point decreases as the number of groups increases but c increases tending to $4-4/n^2$ as $q \to \infty$. Similarly we find the shrinkage factor at $F=c \to 1/2$ as $q \to \infty$. This estimate of $\sigma_b^2=0$ if $F<4-4/n^2$ I find unsatisfactory, especially when it is noted that when experiments are designed to estimate $\sigma_b^2/(\sigma^2+\sigma_b^2)$ efficiently the group size is chosen so that the F value is $2+2\sigma_b^2/\sigma^2$ or approximately 2 if σ_b^2/σ^2 is small (ROBERTSON, 1959).

It is impossible to do similar algebra for LEONARD's (1972) case of binomial data. But numerical examples show that there is a critical value approximately $4q$ and if the χ^2 value is less than this the values shrink to zero.

Going back to linear models, we should note that O'HAGAN (1976), arguing intuitively, suggested that the mode of the posterior distribution of just the variance parameters should be used. Also, to complete the circle, HARVILLE (1974) pointed out that the ML estimation can be thought of as finding the mode of the posterior distribution of the variance parameters and that REML estimation is finding the mode of the posterior distribution of the variance parameters. In both cases using flat priors for α and the variance parameters.

LAIRD (1978) has recently discussed a method that is equivalent to the method of the first paragraph of this section. LAIRD was interested in an $R \times C$ contingency table with a model $\log p_{ij}=r_i+c_j+b_{ij}$ where p_{ij} represents a proportion, and she assumed r_i and c_j had flat vague prior distributions and b_{ij} is normally distributed with variance σ_b^2. Her estimates of r_i and c_j satisfy equations similar to (3.8) with the modification that we now have a log-linear model and an underlying POISSON distribution.

To estimate σ_b^2 LAIRD notes that in a large class of problems (DEMPSTER, RUBIN and LAIRD, 1977) the ML estimates satisfy equations of the form-Expectation of some function given data = Unconditional expectation of function. In her case this equation is $\mathsf{E}(\sum b_{ij}^2 \mid \text{data})=\mathsf{E}(\sum b_{ij}^2)=RC\sigma_b^2$. However the left hand side is complicated to evaluate and she suggests approximating this using the posterior mode of b_{ij} and assuming this mode to be normally distributed. This gives a term $\sum \hat{b}_{ij}^2$ and the variance term contributes a term similar to the W term to arrive at an equation similar to (3.13) to estimate σ_b^2.

We should note that LAIRD was interested in finding a smoother contingency table as a compromise between the observations and a model fitted assuming independence of rows and columns. In some other cases the model might have some physical interpretation or justification. The model used by LAIRD leads to the POISSON-Lognormal distribution. This distribution like the negative binomial, has a variance of the form mean + mean$^2\,\theta$ (where $\theta=\exp(\sigma_b^2)-1$) and has been applied to species-abundance data (BULMER, 1974).

References

ANDERSON, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1**, 135–141.

BULMER, M. G. (1974). On fitting the Poisson Lognormal distribution to Species-Abundance data. *Biometrics* **30**, 101–110.

CURNOW, W. R. N. and SMITH, C. S. (1975). Multifactorial models for familial diseases in man (with Discussion). *J. Roy. Statist. Soc.* A **138**, 131–169.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc.* B **39**, 1–38.

HARVILLE, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* **61**, 383–385.

HARVILLE, D. A. (1977). Maximum likelihood approaches to variance components estimation and to related problems. *J. Amer. Statist. Assoc.* **72**, 320–338.

HAZEL, L. M. (1943). The genetic bases for constructing selection indexes. *Genetics* **28**, 476–490.

HENDERSON, C. R. (1955). Estimation of variance and covariance components. *Biometrics* **9**, 226–252.

HENDERSON, C. R. (1963). Selection index and expected genetic advance. In Statistical Genetics and Plant Breeding (ed. W. D. Hanson and H. F. Robinson), pp. 141–163. Publ. 982. National Academy of Sciences, National Research Council, Washington D. C.

HENDERSON, C. R. (1973). Sire evaluation and genetic trend. In Proc. Anim. Breed. Genet. Symp., Blacksburg, Virginia, 10–41, American Society of Animal Science, Champaign, Illinois.

HENDERSON, C. R., KEMPTHORNE, O., SEARLE, S. R. and VON KROSIGK, C. N. (1959). Estimation of environmental and genetic trends from records subject to culling. *Biometrics* **13**, 192–218.

HEMMERLE, W. J. and HARTLEY, H. O. (1973). Computing maximum likelihood estimates for the mixed A.O.V. model using the W transformation. *Technometrics* **15**, 819–831.

HILL, W. G. and NICHOLAS, F. W. (1974). Estimation of heritability by both regression of offspring on parent and intra-class correlation of sibs in one experiment. *Biometrics* **30**, 447–468.

HILL, W. G. and THOMPSON, R. (1978). Probabilities of non-positive definite between-groups or genetic covariances matrices. *Biometrics* **34**, 429–439.

LAIRD, N. M. (1978). Empirical Bayes methods for two-way contingency tables. *Biometrika* **65**, 581–590.

LEONARD, T. (1972). Bayesian methods for binomial data. Biometrika **51**, 581–589.

LINDLEY, D. V. and SMITH, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *J. Roy. Statist. Soc.* B **34**, 1–41.

NELDER, J. A. (1965). The analysis of randomized experiments with orthogonal structure. *Proc. Roy. Soc.* A **283**, 147–178.

NELDER, J. A. (1968). The combination of information in generally balanced designs. *J. Roy. Statist. Soc.* B **30**, 303–311.

NELDER, J. A. and WEDDERBURN, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc.* A **135**, 370–384.

NOVICK, M. R., JACKSON, P. H. and THAYER, D. T. (1971). Bayesian inference and the classical test theory model reliability and true scores. *Psychometrika* **36**, 207–238.

PATTERSON, H. D. and THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554.

PATTERSON, H. D. and THOMPSON, R. (1975). Maximum likelihood estimation of components of variance. Proceedings of the 8th International Biometric Conference, 197–207.

RAO, C. R. (1973). *Linear Statistical Inference and its Applications* (Second Edition). Wiley, New York, N.Y.

ROBERTSON, A. (1959). Experimental design in the evaluation of genetic parameters. *Biometrics* **15**, 219–226.

ROBERTSON, A. (1962). Weighting in estimation of variance components. *Biometrics* **18**, 413–415.

SMITH, A. F. M. (1973). Bayes estimates in one-way and two-way models. *Biometrika* **60**, 319–329.

THOMPSON, R. (1969). Iterative estimation of variance components for nonorthogonal data. *Biometrics* **25**, 767–773.

THOMPSON, R. (1973). The estimation of variance and covariance components with an application when records are subject to culling. *Biometrics* **29**, 527–550.

THOMPSON, R. (1976a). Design of experiments to estimate heritability when observations are available on parents and offspring. *Biometrics* **32**, 283–304.

THOMPSON, R. (1976b). The estimation of maternal genetic variances. *Biometrics* **32**, 903–918.

THOMPSON, R. (1977). The estimation of heritability with unbalanced data. I. Observations available on parents and offspring. *Biometrics* **33**, 485–495.

THOMPSON, R. (1979). Sire evaluation. *Biometrics* **35**. In press.

THOMPSON, W. A., Jr. (1962). The problem of negative estimates of variance components. *Ann. Math. Statist.* **33**, 273–289.

WILSON, F. S. and HILFERTY, M. M. (1931). The distribution of chi-square. *Proc. Nat. Acad. Sci.*, USA, **17**, 684–688.

YATES, F. and ZACOPANAY, J. (1935). The estimation of the efficiency of sampling with special reference to sampling for yield in cereal experiments. *Journal of Agricultural Science* **25**, 545–577.

YATES, F. (1940). The recovery of inter-block information in balanced incomplete block designs. *Annals of Eugenics* **10**, 317–325.

5

The estimation of variance and covariance components with an application when records are subject to culling

by

R. Thompson

Reprinted from

BIOMETRICS *29*, 527–550
September 1973

THE ESTIMATION OF VARIANCE AND COVARIANCE COMPONENTS WITH AN APPLICATION WHEN RECORDS ARE SUBJECT TO CULLING

R. THOMPSON

A.R.C. Unit of Statistics, University of Edinburgh

SUMMARY

A maximum likelihood (ML) procedure is suggested for the estimation of variance and covariance components. The method consists of maximizing the likelihood, not of all the data, but of a set of selected error contrasts and is developed for the mixed model in the two-way classification. Generalizations when there are more than one random factor are briefly indicated. The results for the two-way classification are then applied to the problem of estimating heritabilities and genetic correlations from parent-offspring data. The results, with minor modifications, are shown to hold even when the parents are selected (or culled) on the basis of the traits measured. A numerical example, with two variates, is given.

1. INTRODUCTION

In this paper we develop a procedure for the estimation of variance and covariance components and fixed effects from multivariate data. We consider this in section 2, in detail, for the general two-way classification when the effects of one factor, for example treatments, are regarded as fixed and the effects of the second factor, such as block or animal effects, are thought of as random. Generalizations to more complicated block structures are briefly indicated.

Variance and covariance components are used by quantitative geneticists as measures of genetic and environmental relationships between two or more characters. Estimation of the components is also required to formulate optimum animal breeding schemes.

Most approaches to the problem of estimating multivariate components deal with the variances and covariances in parallel. In the corresponding univariate case we can estimate variance components by equating various sums of squares to their expectation. Obviously we can extend this process to the multivariate case, estimating covariance components by equating the sums of cross-products to their expectation. This is what Grossman and Gall [1968] and Rohde and Tallis [1969] do.

In the present paper we use instead a ML procedure. This extends the work of Patterson and Thompson [1971] in the univariate situation. Previously, Searle [1956] used the ML method to estimate variance and covariance components for one-way classifications of two variates.

The theory of the method is developed in section 2. In section 3 we discuss how the theory can be implemented. In section 4 we apply our results to the commonly recurring problem of estimating heritabilities and genetic correlations from parent-offspring data.

Hazel [1943] has shown how the genetic correlations can be derived from the sample variances and covariances when the data consist of a complete set of observations on N parent-offspring pairs of individuals.

Frequently, however, records are available for the offspring of $m (m < N)$ parents. For example, in milk yield studies some cows have only male progeny. In this example the sample of parents can be regarded as effectively random. In this case the sample variance-covariance matrix for the m pairs on which we have complete records can be used but this is not efficient since it ignores the information from the $(N - m)$ parents with no recorded offspring. Curnow [1961] pointed out, in a similar univariate case, that the ML method uses this information.

Frequently, however, there is a deliberate selection of parents according to some trait. Thus heifer calves are often retained only from the highest yielding cows in a dairy herd.

We show that for such selected data we can use the same set of estimating equations as when there is no selection, but with a modified information matrix.

Hazel's [1943] formulae give biased estimates of genetic correlations when there is selection, the size of the bias naturally depending on the intensity of selection and the correlation between traits. Van Vleck [1968] gives the results of some Monte Carlo simulations of this method of estimation for varying intensities of truncated selection and different values of the environmental and genetic correlation. He uses his simulations to give regression equations to predict the bias when there is truncation selection on one or two traits.

Brown and Turner [1968] use Tallis's [1961] results on the moment-generating function of the truncated multinormal distribution to discuss the bias in correlations when there is truncation selection in one trait. Their results suggest that Van Vleck's predictive regression equations could be improved. Other workers, for example Purser [1967], have used arguments similar to those of Brown and Turner to correct for biases due to selection.

Reeve [1953] considers a slightly more complicated situation, when measurements on both parents are available and there is assortative mating. He suggests estimating, by weighted least squares, some variance and covariance components from the sample variance-covariance matrix in the parent generation, and some from the parent-offspring regressions. The information on the variances and covariances in the offspring generation does not seem to be utilized.

Since our ML method takes account of the selection there should be little, if any, bias both of the estimates of the variance and covariance components and of the estimates of the fixed effects, for instance, the difference in mean performance between the parent and offspring generation. This

latter point is considered more generally in the univariate case by Henderson *et al.* [1959].

2. THE GENERAL TWO-WAY CLASSIFICATION

We suppose that the two-way classification has t treatments and n units, plots, in b blocks, not necessarily of equal size. Suppose, further, that we have information on p variates and that for each of them the $n \times 1$ vector of yields $\mathbf{y}_j$ $(j = 1, \cdots, p)$ can be represented by the model

$$\mathbf{y}_j = \mathbf{X}\boldsymbol{\alpha}_j + \boldsymbol{\varepsilon}_j \ (j = 1, \cdots, p). \tag{1}$$

Here $\mathbf{X}$ is an $n \times t$ design matrix of rank t, $\boldsymbol{\alpha}_j$ is a $t \times 1$ vector representing the treatment effects on the jth variate. Also the $\boldsymbol{\varepsilon}_j$ are random variables normally distributed with mean zero such that

$$\begin{aligned} E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j') &= \mathbf{Z}\mathbf{Z}'\rho_{ij} + \mathbf{I}_n\Sigma_{ij} (i = 1, \cdots, p; i \geq j) \\ E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j') &= \mathbf{Z}\mathbf{Z}'\rho_{ji} + \mathbf{I}_n\Sigma_{ji} (i = 1, \cdots, p; j > i). \end{aligned} \tag{2}$$

$\mathbf{Z}$ is a $n \times b$ matrix with elements Z_{cd} equal to 1 when unit c is in block d $(c = 1, \cdots, n; d = 1, \cdots, b)$ and equal to zero elsewhere. ρ_{ij} and Σ_{ij} can be thought of as covariances between the ith and jth variates $(i \geq j)$.

In the present paper we assume that the same design and block structure matrices are applicable to all the p variates. Sometimes we require different $\mathbf{X}$ for different variates, e.g. when some treatments are known not to affect certain variates. It also occasionally happens that $\mathbf{Z}$ varies from variate to variate. For example some chemical analyses of grass samples are expensive, both in materials and manpower, and so it is natural to make determinations on bulked samples, although measurements are made on individual plots for such variates as total yield. We shall not consider the additional problems that enter when the pattern of the algebra is destroyed in such ways.

Defining $\mathbf{y}$, $\boldsymbol{\alpha}$, $\boldsymbol{\varepsilon}$ by $R_p[\mathbf{y}] = (\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p)$, $R_p[\boldsymbol{\alpha}] = (\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_p)$, $R_p[\boldsymbol{\varepsilon}] = (\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_p)$ and $\boldsymbol{\Sigma}$ and $\boldsymbol{\varrho}$ as symmetric matrices with i, jth elements Σ_{ij} and ρ_{ij} if $i \geq j$ and Σ_{ji} and ρ_{ji} if $j > i$, (the operator R_p, which transforms a column vector into a p column matrix, is defined in the appendix) then the equations in (1) can be combined to give the compact form

$$\mathbf{y} = (\mathbf{I}_p * \mathbf{X})\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \tag{3}$$

where $\boldsymbol{\varepsilon}$ is a random variable normally distributed with mean zero and variance-covariance matrix given by

$$\mathbf{V} = \boldsymbol{\Sigma} * \mathbf{I}_n + \boldsymbol{\varrho} * \mathbf{Z}\mathbf{Z}'. \tag{4}$$

(The direct product operator $*$ is defined in the appendix. In particular if $p = 2$,

$$\mathbf{I}_p * \mathbf{X} = \begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{X} \end{bmatrix}$$

where $\mathbf{0}$ is $n \times t$ matrix with each element zero, and

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{21} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}$$

where

$$\mathbf{V}_{ij} = \Sigma_{ij}\mathbf{I}_n + \rho_{ij}ZZ')$$

We assume $\boldsymbol{\Sigma}$ is a positive definite matrix and that $\boldsymbol{\varrho}$ is nonsingular. If $\boldsymbol{\varrho}$ is a positive definite matrix then Σ_{ij} and ρ_{ij} can be thought as residual and block covariance components between the ith and jth variates.

The variances of linear combinations of the p variate means for the jth block are functions of $\boldsymbol{\varrho}$ and $\boldsymbol{\Sigma}$ and are positive if the smallest latent root (l_s) of $\boldsymbol{\varrho}\boldsymbol{\Sigma}^{-1}$ is greater than $-1/k_j$, where k_j is the number of plots in the jth block. It is reasonable therefore to impose the condition l_S is greater than $-1/k_{\max}$ where $k_{\max}$ is the largest number of plots in a single block. This, along with $\boldsymbol{\Sigma}$ being nonsingular ensures that $\mathbf{V}$ is nonsingular. When $p = 1$ the condition reduces to $\rho_{11}/\Sigma_{11} > -1/k_{\max}$. Other conditions required for the estimability of $\boldsymbol{\varrho}$ and $\boldsymbol{\Sigma}$ are discussed in section 3.

We need to estimate $\boldsymbol{\alpha}$, $\boldsymbol{\varrho}$, and $\boldsymbol{\Sigma}$. In the univariate case, the reparameterization $\boldsymbol{\gamma} = \boldsymbol{\varrho}\boldsymbol{\Sigma}^{-1}$ is useful and reduces the algebra (Hartley and Rao [1967], Patterson and Thompson [1971]). However, as $\boldsymbol{\varrho}$ and $\boldsymbol{\Sigma}$ are symmetric, we find it simpler to differentiate the likelihood with respect to $\boldsymbol{\varrho}$ and $\boldsymbol{\Sigma}$ than to differentiate with respect to $\boldsymbol{\gamma}$ and $\boldsymbol{\Sigma}$.

We denote estimates of $\boldsymbol{\alpha}$, $\boldsymbol{\varrho}$, and $\boldsymbol{\Sigma}$ by $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\varrho}}$, and $\hat{\boldsymbol{\Sigma}}$. We also use the circumflex to denote functions of $\hat{\boldsymbol{\varrho}}$ and $\hat{\boldsymbol{\Sigma}}$, e.g. $\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{\varrho}}\hat{\boldsymbol{\Sigma}}^{-1}$.

The logarithm of the likelihood of $\mathbf{y}$ is given by

$$L = \text{constant} - \tfrac{1}{2}\log|\mathbf{V}| - (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}). \qquad (5)$$

Following Patterson and Thompson [1971] we divide L into two parts, one part, L', based on contrasts with zero expectation, i.e. error contrasts, the other part, L'', based on contrasts between treatment totals. Arguing that the contrasts in L'' provide no information on error, we estimate $\boldsymbol{\varrho}$ and $\boldsymbol{\Sigma}$ by maximizing L' and estimate $\boldsymbol{\alpha}$ by maximizing L''. L' is the logarithmic likelihood of $\mathbf{S}_p\mathbf{y}$ and L'' the log likelihood of $\mathbf{Q}\mathbf{y}$ where (i) $\mathbf{S}_p$ is a matrix of rank $p(n - t)$ and $\mathbf{Q}$ is a matrix of rank pt (ii) cov $(\mathbf{S}_p\mathbf{y}, \mathbf{Q}\mathbf{y})$ is zero. Suitable matrices for $\mathbf{S}_p$ and $\mathbf{Q}$ are given by

$$\mathbf{S}_p = \mathbf{I}_p * (\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}) = \mathbf{I}_p * \mathbf{S}$$

and

$$\mathbf{Q} = (\mathbf{I}_p * \mathbf{X}')\mathbf{V}^{-1}.$$

We note that $E(\mathbf{S}_p\mathbf{y}) = \mathbf{0}$ and $\mathbf{Q}(\mathbf{I}_p * \mathbf{X})$ is of rank pt so that every linear function of the elements of $\mathbf{Q}\mathbf{y}$ estimates a linear function of the elements of $\boldsymbol{\alpha}$.

Estimates of $\boldsymbol{\varrho}$ and $\boldsymbol{\Sigma}$ are required for the transformation $\mathbf{Q}\mathbf{y}$. Suppose we have estimates $\hat{\boldsymbol{\varrho}}$ and $\hat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\varrho}$ and $\boldsymbol{\Sigma}$. We then estimate $\boldsymbol{\alpha}$ by maximizing L'', the likelihood of $\mathbf{Q}\mathbf{y}$. The estimate $\boldsymbol{\alpha}$ then satisfies the equation

$$\begin{aligned}\alpha &= [(I_p * X')\hat{V}^{-1}(I_p * X)]^{-1}(I_p * X')\hat{V}^{-1}y \\ &= [(I_p * X')\hat{H}^{-1}(I_p * X)]^{-1}(I_p * X')\hat{H}^{-1}y\end{aligned} \tag{6}$$

where

$$\hat{H} = (I_{np} + \hat{\gamma} * ZZ') \quad \text{so} \quad \hat{H}(\hat{\Sigma} * I_p) = \hat{V}.$$

When $p = 2$, equation (6) can be written as

$$\begin{bmatrix}\hat{\alpha}_1 \\ \hat{\alpha}_2\end{bmatrix} = \begin{bmatrix}X'\hat{V}_{11}^{-1}X & X'\hat{V}_{21}^{-1}X \\ X'\hat{V}_{21}^{-1}X & X'\hat{V}_{22}^{-1}X\end{bmatrix}^{-1}\begin{bmatrix}X'(\hat{V}_{11}^{-1}y_1 + \hat{V}_{21}^{-1}y_2) \\ X'(\hat{V}_{21}^{-1}y_1 + \hat{V}_{22}^{-1}y_2)\end{bmatrix}$$

$$= \begin{bmatrix}X'\hat{H}_{11}^{-1}X & X'\hat{H}_{12}^{-1}X \\ X'\hat{H}_{21}^{-1}X & X'\hat{H}_{22}^{-1}X\end{bmatrix}^{-1}\begin{bmatrix}X'(\hat{H}_{11}^{-1}y_1 + \hat{H}_{12}^{-1}y_2) \\ X'(\hat{H}_{21}^{-1}y_1 + \hat{H}_{22}^{-1}y_2)\end{bmatrix}$$

where

$$H = \begin{bmatrix}H_{11} & H_{12} \\ H_{21} & H_{22}\end{bmatrix},$$

$$H_{ij} = I_n + \gamma_{ij}ZZ',$$

$$\begin{bmatrix}V_{11}^{-1} & V_{21}^{-1} \\ V_{21}^{-1} & V_{22}^{-1}\end{bmatrix} = V^{-1} \quad \text{and} \quad \begin{bmatrix}H_{11}^{-1} & H_{12}^{-1} \\ H_{21}^{-1} & H_{22}^{-1}\end{bmatrix} = H^{-1}.$$

An equivalent procedure is to minimize the weighted sum of squares

$$[y - (I_p * X)\alpha]'V^{-1}[y - (I_p * X)\alpha].$$

A convenient form for the inverse of H is

$$H^{-1} = I_{np} - (I_p * Z)(I_p * Z'Z + \gamma^{-1} * I_b)^{-1}(I_p * Z') \tag{7}$$

and this requires only the inverse of b $p \times p$ matrices of the form $I_p k_d + \gamma^{-1}$ where k_d is the number of plots in the dth block ($d = 1, \cdots, b$), many of which will be the same.

The estimation of ϱ and Σ is based on maximizing L' the likelihood of the error contrasts $S_p y$. $S_p = I_p * S$ and S is symmetric, idempotent, of rank $(n - t)$ and independent of ϱ and Σ. The elements of $S_p y$ (or equivalently Sy_j, $j = 1, \cdots, p$) are deviations from treatment means. In the analogous one dimensional case Patterson and Thompson transform Sy into $(n - t)$ orthogonal contrasts $P'Sy$ and set up equations to estimate ϱ and Σ which are functions of these orthogonal contrasts $P'Sy$. In the appendix it is shown how their arguments can be generalized to deal with the multivariate situation using the contrasts $P'Sy_j$ ($j = 1, \cdots, p$).

Patterson and Thompson express their estimating equations in terms arising from modified least squares equations. This approach is usually easier to use since the transformation matrix P' can be difficult to compute. We will therefore follow this approach in presenting the estimating equations for the multivariate case.

We first define $\mathbf{W}$, $\mathbf{U}$, and β, $bp \times bp$, $bp \times bp$, and $bp \times 1$ matrices respectively, as

$$\mathbf{W} = \mathbf{I}_p * \mathbf{Z}'\mathbf{S}\mathbf{Z} + \gamma^{-1} * \mathbf{I}_b , \tag{8}$$

$$\mathbf{U} = (\varrho^{-1} * \mathbf{I}_b)(\mathbf{I}_{bp} - \mathbf{W}^{-1}(\varrho^{-1} * \mathbf{I}_b)), \tag{9}$$

and

$$\beta = \mathbf{W}^{-1}(\mathbf{I}_p * \mathbf{Z}'\mathbf{S})\mathbf{y}. \tag{10}$$

The conditions relating ϱ, Σ, and $k_{\max}$ ensure that W is nonsingular. We might note that α and β satisfy

$$\begin{aligned} (\mathbf{I}_p * \mathbf{X}'\mathbf{X})\alpha + \quad & (\mathbf{I}_p * \mathbf{X}'\mathbf{Z})\beta = (\mathbf{I}_p * \mathbf{X}')\mathbf{y} \\ (\mathbf{I}_p * \mathbf{Z}'\mathbf{X})\alpha + (\mathbf{I}_p * \mathbf{Z}'\mathbf{Z} + \gamma^{-1} * \mathbf{I}_p)\beta & = (\mathbf{I}_p * \mathbf{Z}')\mathbf{y}. \end{aligned} \tag{11}$$

If α is eliminated from equations (11) we get equation (10), if β is eliminated from equations we get equation (6) with $\mathbf{H}^{-1}$ given by equation (7). Just as with α and $\mathbf{y}$ we can define β_i as the jth column of $R_p[\beta]$.

The structure of equations (8), (10), and (11) is probably seen more clearly if we consider the case $p = 2$. Then

$$\mathbf{W} = \begin{bmatrix} \mathbf{Z}'\mathbf{S}\mathbf{Z} + \gamma_{11}^{-1}\mathbf{I}_b & \gamma_{12}^{-1}\mathbf{I}_b \\ \gamma_{21}^{-1}\mathbf{I}_b & \mathbf{Z}'\mathbf{S}\mathbf{Z} + \gamma_{22}^{-1}\mathbf{I}_b \end{bmatrix},$$

$$\begin{aligned} (\mathbf{Z}'\mathbf{S}\mathbf{Z} + \gamma_{11}^{-1}\mathbf{I}_b)\beta_1 + \quad & \gamma_{12}^{-1}\mathbf{I}_b\beta_2 = \mathbf{Z}'\mathbf{S}\mathbf{y}_1 \\ \gamma_{21}^{-1}\mathbf{I}_b\beta_1 + (\mathbf{Z}'\mathbf{S}\mathbf{Z} + \gamma_{22}^{-1}\mathbf{I}_b)\beta_2 & = \mathbf{Z}'\mathbf{S}\mathbf{y}_2 \\ \mathbf{X}'\mathbf{X}\alpha_1 + \quad \mathbf{X}'\mathbf{Z}\beta_1 \quad & = \mathbf{X}'\mathbf{y}_1 \\ \mathbf{X}'\mathbf{X}\alpha_2 + \quad \mathbf{X}'\mathbf{Z}\beta_2 & = \mathbf{X}'\mathbf{y}_2 \\ \mathbf{Z}'\mathbf{X}\alpha_1 + (\mathbf{Z}'\mathbf{Z} + \gamma_{11}^{-1}\mathbf{I}_b)\beta_1 + \quad & \gamma_{12}^{-1}\mathbf{I}_b\beta_2 = \mathbf{Z}'\mathbf{y}_1 \\ \mathbf{Z}'\mathbf{X}\alpha_2 + \quad \gamma_{21}^{-1}\mathbf{I}_b\beta_1 + (\mathbf{Z}'\mathbf{Z} + \gamma_{22}^{-1}\mathbf{I}_b)\beta_2 & = \mathbf{Z}'\mathbf{y}_2 \end{aligned}$$

(where γ_{ij}^{-1} is the i, jth element of γ^{-1}).

We can see then that the only connection between the variates is through the off-diagonal terms of γ^{-1}.

If the off diagonal terms are zero we can split the equation in (11) into p separate parts of the form

$$\begin{aligned} \mathbf{X}'\mathbf{X}\alpha_j + \quad & \mathbf{X}'\mathbf{Z}\beta_j = \mathbf{X}'\mathbf{y}_j \\ \mathbf{Z}'\mathbf{X}\alpha_j + (\mathbf{Z}'\mathbf{Z} + \gamma_{jj}^{-1}\mathbf{I}_b)\beta_j & = \mathbf{Z}'\mathbf{y}_j \end{aligned}$$

and these equations are the basis of Patterson and Thompson's univariate approach. Note that if $\gamma_{jj}^{-1} = 0$ we have the familiar least squares equations for a model assuming fixed treatment and block effects. β_j can be thought of as a vector of modified block effects on the jth variate. Henderson in Henderson *et al.* [1959] has pointed out that in some univariate genetic situations the β_j have some practical interpretation. $\mathbf{W}$ naturally arises when we solve for these "block effects" adjusted for treatments.

We next let $\mathbf{B}$, $\mathbf{A}$, $\mathbf{E}$, and $\mathbf{F}$ be symmetric $p \times p$ matrices such that

$$2\mathbf{B} = \varrho^{-1}R_p[\beta]'R_p[\beta]\varrho^{-1} \tag{12}$$

$$2\mathbf{A} = \mathbf{\Sigma}^{-1}R_p[\mathbf{y}]'\mathbf{S}R_p[\mathbf{y}]\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}R_p[\mathbf{y}]'\mathbf{SZ}R_p[\beta]\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\varrho\mathbf{B} \tag{13}$$

$$2\mathbf{E} = \operatorname{tr}[\mathbf{U}]_p \tag{14}$$

$$2\mathbf{F} = (n - t - b)\mathbf{\Sigma}^{-1} + \operatorname{tr}[(\varrho^{-1} * \mathbf{I}_b)\mathbf{W}^{-1}]_p\,. \tag{15}$$

(The trace operator tr [ ]$_p$ is defined in the appendix.) **B** can be thought of as a weighted sum of squares and cross-products matrix of the "block effects" β. $R_p[\mathbf{y}]'\mathbf{S}R_p[\mathbf{y}]$ is the sum of squares and cross-products matrix of residuals after fitting treatments. $R_p[\mathbf{y}]'\mathbf{SZ}R_p[\beta]$ can be thought of representing the extra sum of squares and cross-products due to the "block effects". Although **A**, **B**, **E**, and **F** are symmetric matrices, the matrices given by $\mathbf{\Sigma}^{-1}R_p[\mathbf{y}]'\mathbf{SZ}R_p[\beta]\mathbf{\Sigma}^{-1}$ and $\mathbf{\Sigma}^{-1}\varrho\mathbf{B}$ are not symmetric but, of course, their sum is symmetric. **F** and **E** are the expectations of **A** and **B** respectively.

The differential of L' with respect to ρ_{ij} and Σ_{ij} can be shown to be $(B_{ij} - E_{ij})$ and $(A_{ij} - F_{ij})$ if $i = j$ and equal to $2(B_{ij} - E_{ij})$ and $2(A_{ij} - F_{ij})$ if $i > j$. More concisely

$$\frac{\partial L'}{\partial \rho_{ij}} = (2 - \delta_{ij})(B_{ij} - E_{ij}) \qquad (i = 1, \cdots, n; j \leq i) \tag{16}$$

$$\frac{\partial L'}{\partial \Sigma_{ij}} = (2 - \delta_{ij})(A_{ij} - F_{ij}) \qquad (i = 1, \cdots, n; j \leq i) \tag{17}$$

(where δ_{ij} is the Kronecker delta).

The estimates of ϱ and $\mathbf{\Sigma}$ maximizing L' therefore satisfy

$$\hat{\mathbf{B}} = \hat{\mathbf{F}} \tag{18}$$

$$\hat{\mathbf{A}} = \hat{\mathbf{E}} \tag{19}$$

and although these are $2p^2$ equations, there are only $p(p + 1)$ independent equations in $p(p + 1)$ unknowns.

The symmetric information matrix (**G**) of $[\varrho]_T$, $[\mathbf{\Sigma}]_T$ can be shown then to be

$$\tfrac{1}{2}(\mathbf{I}_2 * \mathbf{J})\begin{bmatrix} \operatorname{tr}[\mathbf{U}_b^* \mathbf{U}]_p, & \operatorname{tr}[\mathbf{U}_b^*((\varrho^{-1} * \mathbf{I}_b)\mathbf{W}^{-1})]_p, \\ \operatorname{tr}[\mathbf{U}_b^*((\varrho^{-1} * \mathbf{I}_b)\mathbf{W}^{-1})]_p, & (n - t - b)[\mathbf{\Sigma}^{-1} * \mathbf{\Sigma}^{-1}] \\ & + \operatorname{tr}[((\varrho^{-1} * \mathbf{I}_b)\mathbf{W}^{-1})_b^*((\varrho^{-1} * \mathbf{I}_b)\mathbf{W}^{-1})]_p, \end{bmatrix}(\mathbf{I}_2 * \mathbf{J}') \tag{20}$$

where **J** is the Jacobian of the transformation from $\mathbf{\Sigma}_R$ to $[\mathbf{\Sigma}]_T$ ($\mathbf{\Sigma} = R_p[\mathbf{\Sigma}_R]$ and the triangular operator [ ]$_T$ is defined in the appendix).

The structure of (20) becomes clearer if we write down some of the terms for $p = 2$. Then $\mathbf{\Sigma}_R = (\Sigma_{11}, \Sigma_{21}, \Sigma_{21}, \Sigma_{22})'$, $[\mathbf{\Sigma}]_T = (\Sigma_{11}, \Sigma_{21}, \Sigma_{22})'$ and

$$\mathbf{J} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

$\mathbf{U}$ can be split into 4 $b \times b$ matrices such that

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{U}_{21} & \mathbf{U}_{22} \end{bmatrix}$$

and then

$$\text{tr}\,[\mathbf{U}_b^* \mathbf{U}]_4 = \begin{bmatrix} \text{tr}\,(\mathbf{U}_{11}\mathbf{U}_{11}) & \text{tr}\,(\mathbf{U}_{11}\mathbf{U}_{12}) & \text{tr}\,(\mathbf{U}_{12}\mathbf{U}_{11}) & \text{tr}\,(\mathbf{U}_{12}\mathbf{U}_{12}) \\ \text{tr}\,(\mathbf{U}_{11}\mathbf{U}_{21}) & \text{tr}\,(\mathbf{U}_{11}\mathbf{U}_{22}) & \text{tr}\,(\mathbf{U}_{12}\mathbf{U}_{21}) & \text{tr}\,(\mathbf{U}_{12}\mathbf{U}_{22}) \\ \text{tr}\,(\mathbf{U}_{21}\mathbf{U}_{11}) & \text{tr}\,(\mathbf{U}_{21}\mathbf{U}_{12}) & \text{tr}\,(\mathbf{U}_{22}\mathbf{U}_{11}) & \text{tr}\,(\mathbf{U}_{22}\mathbf{U}_{12}) \\ \text{tr}\,(\mathbf{U}_{21}\mathbf{U}_{21}) & \text{tr}\,(\mathbf{U}_{21}\mathbf{U}_{22}) & \text{tr}\,(\mathbf{U}_{22}\mathbf{U}_{21}) & \text{tr}\,(\mathbf{U}_{22}\mathbf{U}_{22}) \end{bmatrix}.$$

Since $\mathbf{U}$ and $(\varrho^{-1} * \mathbf{I}_b)\mathbf{W}^{-1}$ are symmetric, the upper right and lower left quadrants of $\mathbf{G}$ are symmetric. It can be shown that the quadrants of $\mathbf{G}^{-1}$ are also symmetric. This implies that the asymptotic covariance between $\hat{\rho}_{ij}$ and $\hat{\Sigma}_{kl}$ is equal to that between $\hat{\rho}_{kl}$ and $\hat{\Sigma}_{ij}$. This result was noted in a simpler situation by Searle [1956].

In theory, at least, there seems no difficulty in extending these results to the multivariate multicomponent case, essentially extending equations (11) in a similar manner to that in which Patterson and Thompson [1971] dealt with the multicomponent model in the univariate situation. The limiting factor seems rather the number of equations we are able to solve, and this will not be discussed further here.

3. PRACTICAL SOLUTION OF EQUATIONS (18) AND (19)

Equations (18) and (19) can be solved by Fisher's iterative method. We start with preliminary estimates $\hat{\varrho}_0$ and $\hat{\mathbf{\Sigma}}_0$ of ϱ and $\mathbf{\Sigma}$. This gives an estimate of $\hat{\gamma}_0$ which we use to calculate $\mathbf{W}$, $\mathbf{U}$, β, and $\mathbf{B}$, $\mathbf{A}$, $\mathbf{E}$, $\mathbf{F}$, and $\mathbf{G}$ in turn from equations (8) to (10) and (12) to (15) and (20).

We let $\partial L'/\partial[\varrho]_T$ and $\partial L'/\partial[\mathbf{\Sigma}]_T$ be $(p \times (p+1)/2) \times 1$ column vectors, the ith element of these vectors being the differentials of L' with respect to the ith element of $[\varrho]_T$ and $[\mathbf{\Sigma}]_T$ respectively. For instance if $p = 2$,

$$\partial L'/\partial[\varrho]_T = \left(\frac{\partial L'}{\partial \rho_{11}} \frac{\partial L'}{\partial \rho_{21}} \frac{\partial L'}{\partial \rho_{22}}\right)'.$$

$\partial L'/\partial[\hat{\varrho}_0]_T$ and $\partial L'/\partial[\hat{\mathbf{\Sigma}}_0]_T$ can be found using equations (16) and (17).

The closer approximations to $\hat{\rho}$ and $\hat{\Sigma}$ are given by

$$\begin{bmatrix} [\hat{\varrho}]_T \\ [\hat{\mathbf{\Sigma}}]_T \end{bmatrix} = \begin{bmatrix} [\hat{\varrho}_0]_T \\ [\hat{\mathbf{\Sigma}}_0]_T \end{bmatrix} + \hat{G}^{-1} \begin{bmatrix} \partial L'/\partial[\hat{\varrho}_0]_T \\ \partial L'/\partial[\hat{\mathbf{\Sigma}}_0]_T \end{bmatrix}.$$

Sometimes, as in the application discussed later in this paper, simpler formulae can be derived for $\mathbf{B}$, $\mathbf{A}$, $\mathbf{E}$, $\mathbf{F}$, and $\mathbf{G}$ and we do not need to calculate $\mathbf{W}$, $\mathbf{U}$, and β explicitly. α need not be estimated until $\hat{\varrho}$ and $\hat{\mathbf{\Sigma}}$ have been found. A convenient expression for $\hat{\alpha}$ can then be found from equation (11). Then

$$\hat{\alpha}_i = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y}_i - \mathbf{Z}\hat{\beta}_i).$$

When b is large, a more practical solution is to find $\hat{\alpha}$ using (6) and (7). We then find $\hat{\beta}$ from equation (11) as

$$\hat{\beta} = (\mathbf{I}_p * \mathbf{Z}'\mathbf{Z} + \hat{\gamma}^{-1} * \mathbf{I}_p)^{-1}(\mathbf{I}_p * \mathbf{Z}')(\mathbf{y} - \mathbf{I}_p * \mathbf{X})\hat{\alpha}.$$

An alternative expression is useful for $\hat{\mathbf{A}}$:

$$2\hat{\mathbf{A}} = \hat{\Sigma}^{-1}(R_p[\mathbf{y}]'R_p[\mathbf{y}] - R_p[\mathbf{y}]'\mathbf{X}R_p[\hat{\alpha}] - R_p[\mathbf{y}]'\mathbf{Z}R_p[\hat{\beta}])\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}\hat{\varrho}\hat{\beta}.$$

The matrix $\hat{\mathbf{W}}^{-1}$ used in calculating terms in $\mathbf{E}$, $\mathbf{F}$, and the information matrix can be written in terms of $\hat{\mathbf{M}} = (\mathbf{I}_p * \mathbf{Z}'\mathbf{Z} + \hat{\gamma}^{-1} * \mathbf{I}_p)^{-1}$ and

$$[(\mathbf{I}_p * \mathbf{X}')\hat{\mathbf{H}}^{-1}(\mathbf{I}_p * \mathbf{X})]^{-1}$$

already used in the estimation of α. The expression is

$$\hat{\mathbf{W}}^{-1} = \hat{\mathbf{M}} + \hat{\mathbf{M}}(\mathbf{I}_p * \mathbf{Z}'\mathbf{X})[(\mathbf{I}_p * \mathbf{X}')\hat{\mathbf{H}}^{-1}(\mathbf{I}_p * \mathbf{X})]^{-1}(\mathbf{I}_p * \mathbf{X}'\mathbf{Z})\hat{\mathbf{M}}. \quad (21)$$

When $\mathbf{G}$ is singular the solution of (18) and (19) by this iterative method fails. It only seems to breakdown for various fairly obvious pathological cases, for example if $n = t$, then ϱ and Σ are not estimable, or if every block comparison is totally confounded with some treatment comparison, then ϱ is not estimable.

If $\mathbf{X}'\mathbf{X}$ is singular, we can replace $(\mathbf{X}'\mathbf{X})^{-1}$ and $(\mathbf{I}_p * \mathbf{X}')\mathbf{V}^{-1}(\mathbf{I}_p * \mathbf{X})$ throughout by generalized inverses and think of t as the rank of $\mathbf{X}$ rather than the number of treatments. The assumption that ϱ was nonsingular was introduced so that the presentation of the results is in a form directly comparable with previously developed univariate techniques. In fact, we need make no assumption about the singularity of ϱ provided that in equations (12) to (15) and (20) we replace

$$(\varrho^{-1} * \mathbf{I}_b)\mathbf{W}^{-1} \quad \text{by} \quad (\varrho * \mathbf{Z}'\mathbf{S}\mathbf{Z} + \Sigma * \mathbf{I}_b)^{-1} = \mathbf{W}^{*-1},$$

$$\varrho^{-1}R_p[\beta] \quad \text{by} \quad R_p[\beta^*] \quad (\text{where } \beta^* = \mathbf{W}^{*-1}(\mathbf{I}_p * \mathbf{Z}'\mathbf{S})\mathbf{y})$$

and calculate $\mathbf{U}$ as $(\mathbf{I}_p * \mathbf{Z}'\mathbf{S}\mathbf{Z})\mathbf{W}^{*-1}$.

4. APPLICATION TO PARENT-OFFSPRING DATA

We now apply our general results to a particular classification commonly used in genetic studies, namely the classification of pairs of parents and offspring. Suppose we have records on N parents and m offspring, and that each parent has at most one recorded offspring. This section concerns the case when the m animals that have recorded progeny are randomly chosen on the basis of their parent's record.

The parents that have offspring can be numbered as $i = 1, \cdots, m$ and those with no offspring as $i = m + 1, \cdots, N$. The observations for the parents can now be written as

$$y_{ij1}(i = 1, \cdots, N; j = 1, \cdots, p)$$

and those for the offspring as

$$y_{i2}(i = 1, \cdots, m; j = 1, \cdots, p).$$

Let also $\mathbf{y}_{i1}$ $(i = 1, \cdots, N)$ and y_{i2} $(i = 1, \cdots, m)$ be $1 \times p$ row vectors representing the records on the ith parent and offspring respectively.

We assume that the parents records are normally distributed with variance matrix $\boldsymbol{\Sigma} + \boldsymbol{\rho}$, about a mean $\boldsymbol{\mu}_1$, that the offspring records are similarly distributed about $\boldsymbol{\mu}_2$ and that the covariance between parent and offspring records is $\boldsymbol{\rho}$. For instance, if $p = 2$ the variance-covariance matrix of $(y_{i11}, y_{i21}, y_{i12}, y_{i22})$ is

$$\begin{bmatrix} \Sigma_{11} + \rho_{11} & \Sigma_{21} + \rho_{21} & \rho_{11} & \rho_{21} \\ \Sigma_{21} + \rho_{21} & \Sigma_{22} + \rho_{22} & \rho_{21} & \rho_{22} \\ \rho_{11} & \rho_{21} & \Sigma_{11} + \rho_{11} & \Sigma_{21} + \rho_{21} \\ \rho_{21} & \rho_{22} & \Sigma_{21} + \rho_{21} & \Sigma_{22} + \rho_{22} \end{bmatrix}.$$

Curnow [1961] discusses fully, in the corresponding univariate situation, the assumption that the variances in the parent and offspring generating are equal. The tests he suggests for the univariate situation can presumably be generalized to the multivariate case but we will not discuss this further here.

$\boldsymbol{\rho}$ depends on the genetic relationships between the traits and could be characterized in terms of the heritabilities and genetic correlations of the traits. For instance if all the genetic variation is additive the heritability of the ith trait is $2\rho_{ii}/(\Sigma_{ii} + \rho_{ii})$ and the genetic correlation between the ith and jth trait is $\rho_{ij}/(\rho_{ii}\rho_{jj})^{1/2}$ (Falconer [1967]). We let

$$N\bar{\mathbf{y}}_0 = \sum_{i=1}^{N} \mathbf{y}_{i1}$$

$$m\bar{\mathbf{y}}_1 = \sum_{i=1}^{m} \mathbf{y}_{i1}$$

$$m\bar{\mathbf{y}}_2 = \sum_{i=1}^{m} \mathbf{y}_{i2}.$$

$\bar{\mathbf{y}}_0$, $\bar{\mathbf{y}}_1$, and $\bar{\mathbf{y}}_2$ are $1 \times p$ row vectors representing the means of the records of all the parents, of the records of all the parents with recorded offspring, and of the records of the offspring, respectively. We let

$$\mathbf{s}^2 = \sum_{i=1}^{N} \mathbf{y}'_{i1}\mathbf{y}_{i1} - N(\bar{\mathbf{y}}'_0\bar{\mathbf{y}}_0)$$

the corrected sum of squares and cross-products matrix, of size $p \times p$, for all the parents. We can similarly define $\mathbf{M}$ a $2p \times 2p$ matrix as the corrected sum of squares and cross-products matrix of the m $(1 \times 2p)$ row vectors $(\mathbf{y}_{i1}, \mathbf{y}_{i2})$ $(i = 1, \cdots, m)$. We can conveniently split $\mathbf{M}$ into 4 $p \times p$ matrices, i.e.

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix}$$

where $\mathbf{M}_{jk} = \sum_{i=1}^{m} \mathbf{y}'_{ij}\mathbf{y}_{ik} - m(\bar{\mathbf{y}}'_j\bar{\mathbf{y}}_k)$ $(j, k = 1, 2)$. $\mathbf{M}_{11}$ and $\mathbf{M}_{22}$ are the corrected sum of squares and cross-products matrices for those parents with recorded offspring, and for offspring. $\mathbf{M}_{21} = \mathbf{M}'_{12}$ and these matrices contain the corresponding sums of cross-products between parents and offspring.

We show in the appendix that the method of section 2 can be applied in this situation. The resulting estimates of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ satisfy

$$\hat{\boldsymbol{\mu}}_1 = \bar{\mathbf{y}}_0$$

and (22)

$$\hat{\boldsymbol{\mu}}_2 = \bar{\mathbf{y}}_2 - (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_0)(\hat{\boldsymbol{\varrho}} + \hat{\boldsymbol{\Sigma}})^{-1}\hat{\boldsymbol{\varrho}}$$

$(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_0)$ can be thought of as a "selection differential" and $(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_0)(\hat{\boldsymbol{\varrho}} + \hat{\boldsymbol{\Sigma}})^{-1}\hat{\boldsymbol{\varrho}}$ as the "response to selection" (Falconer [1967]). Since in this case parents are selected at random $E(\bar{y}_1) = \bar{y}_0$.

Estimates of $\boldsymbol{\varrho}$ and $\hat{\boldsymbol{\Sigma}}$ satisfy

$$\hat{\mathbf{A}} = \hat{\mathbf{F}} \tag{23}$$

$$\hat{\mathbf{B}} = \hat{\mathbf{E}} \tag{24}$$

where

$$\begin{aligned} 2\mathbf{A} = {} & (\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1}(\mathbf{s}^2 - \mathbf{M}_{11})(\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} \\ & + \tfrac{1}{2}(2\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1}(\mathbf{M}_{11} + \mathbf{M}_{12} + \mathbf{M}_{21} + \mathbf{M}_{22})(2\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} \\ & + \tfrac{1}{2}\boldsymbol{\Sigma}^{-1}(\mathbf{M}_{11} - \mathbf{M}_{12} - \mathbf{M}_{21} + \mathbf{M}_{22})\boldsymbol{\Sigma}^{-1} \end{aligned} \tag{25}$$

$$2\mathbf{F} = (N - m)(\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} + (m - 1)(2\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} + (m - 1)\boldsymbol{\Sigma}^{-1} \tag{26}$$

$$\begin{aligned} 2\mathbf{B} = {} & (\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1}(\mathbf{s}^2 - \mathbf{M}_{11})(\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} \\ & + (2\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1}(\mathbf{M}_{11} + \mathbf{M}_{12} + \mathbf{M}_{21} + \mathbf{M}_{22})(2\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} \end{aligned} \tag{27}$$

$$2\mathbf{E} = (N - m)(\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} + 2(m - 1)(2\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} \tag{28}$$

$(\mathbf{M}_{11} + \mathbf{M}_{12} + \mathbf{M}_{21} + \mathbf{M}_{22})$ and $(\mathbf{M}_{11} - \mathbf{M}_{12} - \mathbf{M}_{21} + \mathbf{M}_{22})$ can be thought of as the sum of squares and cross-products matrices of the m vectors $(\mathbf{y}_{i1} + \mathbf{y}_{i2})$ and $(\mathbf{y}_{i1} - \mathbf{y}_{i2})$ $(i = 1, \cdots, m)$, respectively. The information matrix (G) of $[\hat{\boldsymbol{\varrho}}]_T$, $[\hat{\boldsymbol{\Sigma}}]_T$ can be found to be

$$\tfrac{1}{2}(\mathbf{I}_2 * \mathbf{J})\begin{bmatrix} \mathbf{J}_{11} & \mathbf{J}_{21} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{bmatrix}(\mathbf{I}_2 * \mathbf{J}')$$

where

$$\begin{aligned} \mathbf{J}_{11} &= (N - m)(\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} * (\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} + 4(m - 1)(2\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} * (2\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} \\ \mathbf{J}_{21} &= (N - m)(\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} * (\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} + 2(m - 1)(2\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} * (2\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} \\ \mathbf{J}_{22} &= (N - m)(\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} * (\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} + (m - 1)(2\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} * (2\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} \\ &\quad + (m - 1)\boldsymbol{\Sigma}^{-1} * \boldsymbol{\Sigma}^{-1}. \end{aligned} \tag{29}$$

Equations (23) and (24) can easily be derived using expressions (25) to (28). In practical examples solutions to equations (23) and (24) have been found quite quickly using the iterative method outlined in section 3.

5. MODIFICATIONS WHEN RECORDS ARE SUBJECT TO CULLING

Sometimes the m parents with offspring are not randomly selected. We discuss in this section the modifications necessary when the probability of an animal having a recorded offspring depends solely on its own record. Sometimes in choosing parents of the next generation information on relatives, such as mothers or half-sisters, could be utilized. A more general model could be suggested to take account of such selection but we do not consider this any further in this paper.

As Van Vleck [1968] mentions there are then three sets of variances and covariances involved (1) the variances and covariances in the parental generation before selection (2) the variances and covariances in the parental generation after selection (3) the variances and covariances in the offspring generation. We estimate the variance and covariance components in the parental generation before selection. The second and third sets of variances and covariances can be expressed in terms of the first set and the selection scheme employed, at least for the simple model of inheritance used in this paper.

By following the approach of Kempthorne and Von Krosigk (in Henderson *et al.* [1959]) and of Curnow [1961], we show in the appendix that the likelihood for this case is a factorization of the likelihood used in section 4. So equations (22) to (28) can be used as a basis to estimate $\mathbf{\mu}_1$, $\mathbf{\mu}_2$, ϱ, and $\mathbf{\Sigma}$ even if the parents that have recorded offspring are selected.

However the expectations of the moments, $\mathbf{M}_{11}$, $\mathbf{M}_{12}$, $\mathbf{M}_{21}$, and $\mathbf{M}_{22}$, used in calculating the information matrix do depend on the selection scheme and we discuss the changes in the appendix.

We will denote the modified information matrix of $[\hat{\varrho}]_T$, $[\hat{\mathbf{\Sigma}}]_T$ by

$$\tfrac{1}{2}(\mathbf{I}_2 * \mathbf{J})\begin{bmatrix}\mathbf{J}_{11}^* & \mathbf{J}_{21}^* \\ \mathbf{J}_{21}^* & \mathbf{J}_{22}^*\end{bmatrix}(\mathbf{I}_2 * \mathbf{J}')$$

where

$$\begin{aligned}
\mathbf{J}_{11}^* &= \mathbf{J}_{11} + [4(2\varrho + \mathbf{\Sigma})^{-1} - 2(\varrho + \mathbf{\Sigma})^{-1}]*\mathbf{M}_{11}^* \\
\mathbf{J}_{21}^* &= \mathbf{J}_{21} + [2(2\varrho + \mathbf{\Sigma})^{-1} - 2(\varrho + \mathbf{\Sigma})^{-1}]*\mathbf{M}_{11}^* \\
\text{and} \quad \mathbf{J}_{22}^* &= \mathbf{J}_{22} + [\mathbf{\Sigma}^{-1} + (2\varrho + \mathbf{\Sigma})^{-1} - 2(\varrho + \mathbf{\Sigma})^{-1}]*\mathbf{M}_{11}^* \\
\mathbf{M}_{11}^* &= (\mathbf{\Sigma} + \varrho)^{-1}[\mathbf{M}_{11} - (m-1)(\mathbf{\Sigma} + \varrho)](\mathbf{\Sigma} + \varrho)^{-1}.
\end{aligned} \tag{30}$$

$\mathbf{M}_{11}^*$ is one measure of the selection employed. If parents with recorded offspring are randomly selected then obviously the expected value of $\mathbf{M}_{11}^* = \mathbf{0}$ and equations (30) reduce to equations (29).

The upper right and lower left quadrants of the information matrix are again symmetric, but in this case the upper right and lower left quadrants of the inverse need not be symmetric.

6. COMPARISON WITH FULL MAXIMUM LIKELIHOOD

As in the univariate case considered by Patterson and Thompson, there is a simple relationship between the modified ML procedure and the full ML method. Both methods use the same equations to estimate $\boldsymbol{\alpha}$ once $\boldsymbol{\varrho}$ and $\boldsymbol{\Sigma}$ are found. In the modified ML method $\boldsymbol{\varrho}$ and $\boldsymbol{\Sigma}$ are estimated by essentially equating $\mathbf{A}$ and $\mathbf{B}$ (equations (25) and (27)) to their expectation. Whilst in the full ML method the same numerical functions $\mathbf{A}$ and $\mathbf{B}$ are used, but they are equated to their expectation assuming $\boldsymbol{\alpha}$ is fixed. There can be an appreciable difference between the two methods if the number of treatments, t, is large compared with the error degrees of freedom, $n - t$.

For the application in section 4 and 5 it can be shown that the full ML estimates of $\boldsymbol{\varrho}$ and $\boldsymbol{\Sigma}$ satisfy $\hat{\mathbf{A}} = \hat{\mathbf{F}}$, $\hat{\mathbf{B}} = \hat{\mathbf{E}}$ with $\mathbf{A}$, $\mathbf{F}$, $\mathbf{B}$, and $\mathbf{E}$ as given in equations (25) to (28), but with $N + 1$ and $m + 1$ replacing N and m in equations (26) and (28). Obviously in this application there is a trivial difference between the two sets of estimates if m is large. For the univariate case, these full ML estimating equations reduce to those given by Curnow [1961]. He also considers the bias in $\rho_{11}/(\rho_{11} + \Sigma_{11})$ and if we follow his arguments we can show that $\hat{\rho}_{11}/(\hat{\rho}_{11} + \hat{\Sigma}_{11})$, the estimate given by the modified ML procedure is asymptotically unbiased (to order $1/m$).

7. NUMERICAL EXAMPLE

In this section the method developed in sections 4 and 5 for parent-offspring data is illustrated, for the case $p = 2$, using generated data. 1000 vectors $(y_{i11}, y_{i21}, y_{i12}, y_{i22})$ $(i = 1, \cdots, 1000)$ were found, using a pseudo-random number generator, from a multivariate normal distribution with means 0 and variance-covariance matrix

$$\begin{bmatrix} 1.0 & 0.5 & 0.25 & 0.125 \\ 0.5 & 1.0 & 0.125 & 0.25 \\ 0.25 & 0.125 & 1.0 & 0.5 \\ 0.125 & 0.25 & 0.5 & 1.0 \end{bmatrix}.$$

We let (y_{i11}, y_{i21}) and (y_{i12}, y_{i22}) represent records on the parent and offspring respectively. In the notation of section 4 the variance-covariance matrix can be written as

$$\begin{bmatrix} \boldsymbol{\Sigma} + \boldsymbol{\varrho} & \boldsymbol{\varrho} \\ \boldsymbol{\varrho} & \boldsymbol{\Sigma} + \boldsymbol{\varrho} \end{bmatrix}$$

where

$$\varrho = \tfrac{1}{4}(\boldsymbol{\Sigma} + \varrho) \quad \text{and} \quad (\boldsymbol{\Sigma} + \varrho) = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}.$$

The means for the parents were

$$(-0.0016 - 0.0607) = \bar{y}_0$$

and the corrected sum of squares and cross products matrix for the 1000 (y_{i11}, y_{i21}) pairs was

$$\begin{bmatrix} 1022.69 & 523.26 \\ 523.26 & 997.76 \end{bmatrix} = s^2.$$

To include the selection of parents into this example only those records of offspring (y_{i12}, y_{i22}) such that y_{i11} was greater than zero were used. There were 501 such vectors and the means of the selected parents and the offspring were

$$(0.8188\ 0.3796) = \bar{y}_1 \quad \text{and} \quad (0.2495\ 0.1172) = \bar{y}_2,$$

and the corrected sum of squares and cross products matrix for the 501 vectors $(y_{i11}, y_{i12}, y_{i21}, y_{i22})$ representing the records of the selected parents and their offspring was

$$\begin{bmatrix} 175.12 & 92.57 & 36.85 & 30.06 \\ 92.57 & 446.72 & 3.18 & 95.70 \\ 36.85 & 3.18 & 466.72 & 207.43 \\ 30.06 & 95.70 & 207.43 & 471.82 \end{bmatrix} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix}.$$

We require estimates of ϱ and $\boldsymbol{\Sigma}$ to initiate the iterative procedure for solving equations (18) and (19). Natural estimates for $\rho_{11}/(\rho_{11} + \Sigma_{11})$ and $\rho_{22}/(\rho_{22} + \Sigma_{22})$ are $36.85/175.12 = 0.210$ and $95.70/446.72 = 0.214$ given by the regression of offspring records on their parents. Since there has been selection of parents only on the first trait in this example, this gives an unbiased estimate of $\rho_{11}/(\rho_{11} + \Sigma_{11})$. If such a selection scheme is followed, usually the estimate of $\rho_{22}/(\rho_{22} + \Sigma_{22})$ will be biased, but for the particular values of ϱ and $\boldsymbol{\Sigma}$ used in this example, it can be shown that the estimate is unbiased. The genetic correlation $(\rho_{21}/(\rho_{11}\rho_{22})^{1/2})$ between the two variates can be estimated using the terms of $\mathbf{M}_{21}$ as (Van Vleck [1968])

$$\tfrac{1}{2}(30.06 + 3.18)/(36.85 \times 95.70)^{1/2} = 0.280.$$

For this selection scheme Brown and Turner's [1968] theoretical results predict a bias of about -0.17 and Van Vleck's [1968] empirical results predict a bias of from -0.14 to -0.17.

Using these estimates of functions of ϱ and $\boldsymbol{\Sigma}$ and the fact that s^2 gives an estimate of $999(\boldsymbol{\Sigma} + \varrho)$ we arrive at initial estimates of ϱ and $\boldsymbol{\Sigma}$,

$$\varrho_0 = \begin{bmatrix} 0.215 & 0.060 \\ 0.060 & 0.214 \end{bmatrix} \qquad \Sigma_0 = \begin{bmatrix} 0.809 & 0.464 \\ 0.464 & 0.785 \end{bmatrix}$$

$$(0.215 = (0.210 \times 1022.59)/999 \text{ etc.}).$$

Using these initial estimates the iterative procedure converges satisfactorily in three iterations. To illustrate the method we will use the estimates

$$\varrho_0 = \begin{bmatrix} 0.215 & 0.000 \\ 0.000 & 0.214 \end{bmatrix} \qquad \Sigma_0 = \begin{bmatrix} 0.809 & 0.000 \\ 0.000 & 0.785 \end{bmatrix}$$

since this makes the matrix arithmetic easier to follow.

We first calculate $\mathbf{s}^2 - \mathbf{M}_{11}$, $\mathbf{M}_{11} + \mathbf{M}_{12} + \mathbf{M}_{21} + \mathbf{M}_{22}$, and $\mathbf{M}_{11} - \mathbf{M}_{12} - \mathbf{M}_{21} + \mathbf{M}_{22}$ as these are the basic matrices in the formation **A** and **B**.

$$\mathbf{s}^2 - \mathbf{M}_{11} = \begin{bmatrix} 847.57 & 430.69 \\ 430.69 & 551.04 \end{bmatrix}$$

$$\mathbf{M}_{11} + \mathbf{M}_{12} + \mathbf{M}_{21} + \mathbf{M}_{22} = \begin{bmatrix} 715.54 & 333.24 \\ 333.24 & 1110.20 \end{bmatrix}$$

$$\mathbf{M}_{11} - \mathbf{M}_{12} - \mathbf{M}_{21} + \mathbf{M}_{22} = \begin{bmatrix} 568.14 & 266.76 \\ 266.76 & 727.40 \end{bmatrix}.$$

Then

$$(\varrho_0 + \Sigma_0)^{-1}(\mathbf{s}^2 - \mathbf{M}_{11})(\varrho_0 + \Sigma_0)^{-1} = \begin{bmatrix} 808.3057 & 421.0167 \\ 421.0167 & 551.8832 \end{bmatrix}$$

$$(2\varrho_0 + \Sigma_0)^{-1}(\mathbf{M}_{11} + \mathbf{M}_{12} + \mathbf{M}_{21} + \mathbf{M}_{22})(2\varrho_0 + \Sigma_0)^{-1} = \begin{bmatrix} 466.1131 & 221.7303 \\ 221.7303 & 754.5354 \end{bmatrix}$$

$$\Sigma_0^{-1}(\mathbf{M}_{11} - \mathbf{M}_{12} - \mathbf{M}_{21} - \mathbf{M}_{22})\Sigma_0^{-1} = \begin{bmatrix} 868.0771 & 420.0515 \\ 420.0515 & 1180.4130 \end{bmatrix}.$$

So **A** and **B** can be found from equations (25) and (27) as to be given by

$$2\mathbf{A} = \begin{bmatrix} 1475.4008 & 741.9076 \\ 741.9076 & 1519.3574 \end{bmatrix} \qquad 2\mathbf{B} = \begin{bmatrix} 1274.4188 & 642.7470 \\ 642.7470 & 1306.4186 \end{bmatrix}.$$

Since $N = 1000$ and $m = 501$, we find **F** and **E** (equations (26) and (28)) to be

$$2\mathbf{F} = \begin{bmatrix} 1508.9029 & 0.0000 \\ 0.0000 & 1548.6433 \end{bmatrix} \qquad 2\mathbf{E} = \begin{bmatrix} 1294.4072 & 0.0000 \\ 0.0000 & 1323.9018 \end{bmatrix}.$$

If parents are selected at random, the information matrix is given by equations (29). For instance

$$\mathbf{J}_{11} = \begin{bmatrix} 1778.7124 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 1818.5468 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 1818.5468 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 1859.2778 \end{bmatrix}.$$

However, since parents are selected, we use equations (20) to give the information matrix. $\mathbf{M}_{11}^{*}$, a measure of the selection scheme is found to be

$$\mathbf{M}_{11}^{*} = \begin{bmatrix} -321.2738 & 90.4909 \\ 90.4909 & -52.6252 \end{bmatrix}$$

and hence

$$\mathbf{J}_{11}^{*} = \begin{bmatrix} 1368.9967 & 115.4017 & 0.0000 & 0.0000 \\ 115.4017 & 1751.4347 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 1402.3021 & 117.2407 \\ 0.0000 & 0.0000 & 117.2407 & 1741.0962 \end{bmatrix}$$

and twice the information matrix is

$$\begin{bmatrix} 1368.9967 & 115.4017 & 0.0000 & 1236.1841 & -30.6692 & 0.0000 \\ 115.4017 & 3153.7368 & 117.2406 & -30.6692 & 2437.6481 & -31.9611 \\ 0.0000 & 117.2406 & 1791.0962 & 0.0000 & -31.9611 & 1266.6427 \\ 1236.1841 & -30.6692 & 0.0000 & 1536.6172 & 8.1506 & 0.0000 \\ -30.6692 & 2437.6481 & -31.9611 & 8.1506 & 3179.9300 & 8.7130 \\ 0.0000 & -31.9611 & 1266.6427 & 0.0000 & 8.7130 & 1646.1440 \end{bmatrix}.$$

The differential of L' with respect to ρ_{11} , ρ_{21} , ρ_{22} , Σ_{11} , Σ_{21} , and Σ_{22} can be found from $\mathbf{B} - \mathbf{E}$ and $\mathbf{A} - \mathbf{F}$, using equations (18) and (19) to be

$$(-9.9942 \quad 642.7470 \quad -8.7416 \quad -16.2511 \quad 741.4832 \quad -14.6430)' = \begin{bmatrix} \partial L'/\partial[\boldsymbol{\varrho}_0]_T \\ \partial L'/\partial[\boldsymbol{\Sigma}_0]_T \end{bmatrix}.$$

Hence premultiplying this vector by the inverse of the information matrix we find better estimates of $\boldsymbol{\varrho}$ and $\boldsymbol{\Sigma}$ are given by

$$\begin{bmatrix} \hat{\rho}_{11} \\ \hat{\rho}_{21} \\ \hat{\rho}_{22} \\ \hat{\Sigma}_{11} \\ \hat{\Sigma}_{21} \\ \hat{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} 0.215 \\ 0.000 \\ 0.214 \\ 0.809 \\ 0.000 \\ 0.785 \end{bmatrix} + \begin{bmatrix} 0.015 \\ 0.112 \\ 0.003 \\ -0.033 \\ 0.381 \\ -0.020 \end{bmatrix}$$

so

$$\hat{\varrho} = \begin{bmatrix} 0.230 & 0.112 \\ 0.112 & 0.217 \end{bmatrix} \quad \text{and} \quad \hat{\Sigma} = \begin{bmatrix} 0.776 & 0.381 \\ 0.381 & 0.765 \end{bmatrix}.$$

After another three cycles the estimates were found to be

$$\hat{\varrho} = \begin{bmatrix} 0.236 & 0.121 \\ 0.121 & 0.222 \end{bmatrix} \quad \text{and} \quad \hat{\Sigma} = \begin{bmatrix} 0.771 & 0.374 \\ 0.374 & 0.761 \end{bmatrix},$$

the changes in the last cycle being of the order 0.0001.

The estimates of μ_1 and μ_2 are then (from (22))

$$\begin{aligned} \hat{\mu}_1 &= (-0.0016 \quad -0.0607) \\ \hat{\mu}_2 &= (0.2495 \quad 0.1172) - (0.1925 \quad 0.1067) \\ &= (0.0570 \quad 0.0105). \end{aligned}$$

The asymptotic variance-covariance matrix ($\mathbf{G}^{-1}$) of ($\hat{\rho}_{11}$, $\hat{\rho}_{21}$, $\hat{\rho}_{22}$, $\hat{\Sigma}_{11}$, $\hat{\Sigma}_{21}$, $\hat{\Sigma}_{22}$) was found to be ($\times 10^6$)

$$\begin{bmatrix} 4930 & 2375 & 1144 & -3841 & -1836 & -877 \\ 2375 & 2138 & 1516 & -1841 & -1611 & -1122 \\ 1144 & 1516 & 2211 & -882 & -1126 & -1585 \\ -3841 & -1841 & -882 & 4226 & 2027 & 972 \\ -1836 & -1611 & -1126 & 2027 & 1961 & 1419 \\ -877 & -1122 & -1585 & 972 & 1419 & 2299 \end{bmatrix}.$$

The upper right and lower left quadrants are almost symmetric. The final estimate of ϱ and Σ agree quite satisfactorily with the values of ϱ and Σ used to generate the data. There is relatively little change from the initial to the final estimates of the variances but, as might be expected considering the bias in the initial estimates, quite a change in the covariances.

It can be seen that the variances of $\hat{\rho}_{11}$ and $\hat{\Sigma}_{11}$ are larger than those of $\hat{\rho}_{22}$ and $\hat{\Sigma}_{22}$, this might be expected since parents with offspring were chosen using a truncation selection scheme on the first variate. If parents had been selected at random the variances of $\hat{\rho}_{11}$ and $\hat{\rho}_{22}$, and $\hat{\Sigma}_{11}$ and $\hat{\Sigma}_{22}$, would have been equal and the estimates of ρ_{11} and Σ_{11} would have been more precise than in this example. It is interesting to note that if we use $\mathbf{G}$ as defined in (29), that is assuming the parents are selected randomly, in the iterative procedure we arrive at the same solution for $\hat{\varrho}$ and $\hat{\Sigma}$ but need twice as many iterations.

L'ESTIMATEUR DES COMPOSANTES DE LA VARIANCE ET DE LA COVARIANCE AVEC APPLICATION AU CAS OÙ LES ÉLÉMENTS ETUDIÉS SONT SUJETS À SÉLECTION

RESUME

On suggère un procédé du maximum de vraisemblance pour l'estimation des composantes de la variance et de la covariance. La méthode consiste à maximiser la vraisemblance,

non de toutes les données, mais d'un ensemble choisi de contrastes d'erreur; elle est appliquée au modèle mixte à deux facteurs de classification. Des généralisations sont brièvement indiquées pour le cas où il y a plus d'un facteur aléatoire. Les résultats pour deux facteurs de classification sont ensuite appliqués au problème d'estimation d'héritabilités et de corrélations génétiques à partir de données parents-descendants. On montre que ces résultats, aprés modifications mineures, restent valables même quand les parents sont sélectionnés à partir des caractères measurés. Un exemple numérique à deux variables est donné.

REFERENCES

Brown, G. H. and Turner, H. N. [1968]. Response to selection in Australian Merino sheep. II Estimates of phenotypic and genetic parameters for some production traits on Merino ewes and an analysis of the possible effects of selection on them. *Aust. J. of Agric. Research 19*, 303–22 (Corrigendum, *21*, 182).

Curnow, R. N. [1961]. The estimation of repeatability and heritability from records subject to culling. *Biometrics 17*, 553–66.

Falconer, D. S. [1967]. *Introduction to Quantitative Genetics*. Oliver and Boyd, Edinburgh.

Grossman, M. and Gall, G. A. E. [1968]. Covariance analysis with unequal subclass numbers: component estimation in quantitative genetics. *Biometrics 24*, 49–59.

Hartley, H. O. and Rao, J. N. K. [1967]. Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika 54*, 93–108.

Hazel, L. N. [1943]. The genetic basis for constructing selection indices. *Genetics 28*, 476–90.

Henderson, C. R., Kempthorne, O., Searle, S. R. and Von Krosigk, C. N. [1959]. Estimation of environomental and genetic trends from records subject to culling. *Biometrics 15*, 192–218.

Patterson, H. D. and Thompson, R. [1971]. Recovery of inter-block information when block sizes are unequal. *Biometrika 58*, 545–54.

Purser, A. F. [1967]. Selection experiments with hill sheep. Report of proceedings and invited papers of the 9th International Congress of Animal Production. 32–40, Oliver and Boyd, Edinburgh.

Reeve, E. C. R. [1953]. Studies in quantitative inheritance. III Heritability and genetic correlation in progeny tests using different mating systems. *J. Genet. 51*, 520–42.

Rohde, C. A. and Tallis, G. M. [1969]. Exact first- and second-order moments of estimates of components of covariance. *Biometrika 56*, 517–26.

Searle, S. R. [1956]. Matrix methods in variance and covariance component analysis. *Ann. Math. Statist. 27*, 737–48.

Searle, S. R. [1966]. *Matrix Algebra for the Biological Sciences*. Wiley, New York.

Tallis, G. M. [1961]. The moment generating function of the truncated multi-normal distribution. *J. R. Statist. Soc. B23*, 223–9.

Van Vleck, L. D. [1968]. Selection bias in estimation of the genetic correlation. *Biometrics 24*, 951–62.

APPENDIX

Definition of matrix operators

We define here the four matrix operators that simplify the presentation of the results.

(i) The direct product of order p.

If **A** is a $m \times n$ matrix and **B** is a $r \times s$ matrix then the direct product of order p of **A** and **B** exists for any positive integer p for which p divides each

of m, n, r, and s and is a $(mr/p) \times (ns/p)$ matrix. The product is denoted by $\mathbf{A}^*_p\mathbf{B}$ and are, with $t = m/p$, $u = n/p$, $v = r/p$, $w = s/p$, given by

$$\mathbf{A}^*_p\mathbf{B} = \left[\begin{array}{ccc|c|ccc} \mathbf{A}_{11}\mathbf{B}_{11} & \cdots & \mathbf{A}_{11}\mathbf{B}_{1w} & & \mathbf{A}_{1u}\mathbf{B}_{11} & \cdots & \mathbf{A}_{1u}\mathbf{B}_{1w} \\ \cdot & \cdots & \cdot & & \cdot & \cdots & \cdot \\ \mathbf{A}_{11}\mathbf{B}_{v1} & \cdots & \mathbf{A}_{11}\mathbf{B}_{vw} & & \mathbf{A}_{1u}\mathbf{B}_{v1} & \cdots & \mathbf{A}_{1u}\mathbf{B}_{vw} \\ \hline & & & & & & \\ \hline \mathbf{A}_{t1}\mathbf{B}_{11} & \cdots & \mathbf{A}_{t1}\mathbf{B}_{1w} & & \mathbf{A}_{tu}\mathbf{B}_{11} & \cdots & \mathbf{A}_{tu}\mathbf{B}_{1w} \\ \cdot & \cdots & \cdot & & \cdot & \cdots & \cdot \\ \mathbf{A}_{t1}\mathbf{B}_{v1} & \cdots & \mathbf{A}_{t1}\mathbf{B}_{vw} & & \mathbf{A}_{tu}\mathbf{B}_{v1} & \cdots & \mathbf{A}_{tu}\mathbf{B}_{vw} \end{array}\right]$$

where $\mathbf{A}_{ij}$ and $\mathbf{B}_{kl}$ are the i, jth and k, lth submatrices of size $p \times p$ of $\mathbf{A}$ and $\mathbf{B}$. When $p = 1$ we have the more usual direct product (Searle [1966]). Our definition is more general and allows us to multiply submatrices together. Since we mostly use the direct product of order 1 we will write, for typographical convenience, $\mathbf{A}^*_1\mathbf{B}$ as $\mathbf{A}^*\mathbf{B}$. We continually use the result that $(\mathbf{A}^*\mathbf{B})(\mathbf{C}^*\mathbf{D}) = \mathbf{A}\mathbf{C}^*\mathbf{B}\mathbf{D}$, if the matrix products exist.

(ii) The general trace of order p.

The general trace of order p of a $mp \times mp$ matrix $\mathbf{C}$ is defined as a $p \times p$ matrix $\mathbf{D}$ and the general trace of $\mathbf{C}$ is denoted by $\text{tr } [\mathbf{C}]_p = \mathbf{D}$. The i, jth element of $\mathbf{D}$ is the trace (i.e. sum of diagonal elements) of $\mathbf{C}_{ij}$, the i, jth submatrix of size $m \times m$ of $\mathbf{C}$. Once again we omit the suffix when $p = 1$, i.e. we write $\text{tr } [\mathbf{C}]_1$ as $\text{tr } [\mathbf{C}]$.

(iii) A vector to matrix operator.

We find it useful to define an operator that rearranges the element of a column vector $\mathbf{y}$, $a(n \times 1)$ vector, and puts them into a matrix $\mathbf{Y}$ of size $(n/p) \times p$. The operator $R_p[\mathbf{y}]$ only exists for any p that is a factor of n. The jth column of $\mathbf{Y}$ contains elements $1 + n(j-1)$, $2 + n(j-1)$, $\cdots$, nj of $\mathbf{y}(j = 1, 2, \cdots, p)$, i.e.

$$R_p[\mathbf{y}] = \mathbf{Y} = \begin{bmatrix} y_1 & y_{1+n} & \cdots & y_{(p-1)n+1} \\ y_2 & y_{2+n} & \cdots & y_{(p-1)n+2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ y_n & y_{2n} & \cdots & y_{pn} \end{bmatrix}.$$

(iv) A matrix to vector operator.

Often in multivariate work one is interested in symmetric matrices. One can then either estimate the lower or upper triangle of these matrices. In this paper we arbitrarily work with the lower triangle, and it is conveneint to define an operation that takes the lower triangle of $\mathbf{A}$, a $p \times p$ symmetric matrix, and puts the elements in a column vector $\mathbf{B}$ of size $(p(p+1)/2) \times 1$. That is, the $i(i-1)/2 + j$th element of $\mathbf{B}$ contains element (i, j) of $\mathbf{A}(i = 1, \cdots, p; j \leq i)$. We denote this operator by $[\mathbf{A}]_T = \mathbf{B}$ and so $\mathbf{B}' = [A_{11}, A_{21}, A_{22}, \cdots, A_{p(p-1)}, A_{pp}]$.

The estimation of ϱ *and* Σ

We estimate ϱ and Σ by maximizing L' the log likelihood of the contrasts $\mathbf{S}_p y$. Since $\mathbf{S}_\nu \mathbf{V} \mathbf{S}_\nu$, the variance matrix of $\mathbf{S}_\nu \mathbf{y}$, is singular, we express L' in terms of a generalized inverse $(\mathbf{S}_p \mathbf{V} \mathbf{S}_p)^{-g}$ of $\mathbf{S}_p \mathbf{V} \mathbf{S}_p$ using the latent roots of $\mathbf{SZZ'S}$.

For this purpose we use an $n \times (N - t)$ matrix P whose columns are orthogonal vectors of both $\mathbf{S}$ and $\mathbf{SZZ'S}$. Then

(i) $\mathbf{P'P} = \mathbf{I}$

(ii) $\mathbf{PP'} = \mathbf{S}$

(iii) $\mathbf{P'ZZP} = \text{diag}\ (\lambda_s)$, a diagonal matrix with elements λ_s , $s = 1, \cdots, n - t$.

It follows from (i) and (ii) that the likelihood of $\mathbf{S}\mathbf{y}_i$ is also the likelihood of $\mathbf{P'}\mathbf{y}_i (j = 1, \cdots, p)$.

Since

$$\mathbf{S}_p \mathbf{V} \mathbf{S}_p = \mathbf{P}_p(\varrho^* \text{ diag } (\lambda_s) + \hat{\Sigma}^* \mathbf{I}_p)\mathbf{P}'_p$$

where $\mathbf{P}_p = \mathbf{I}_p * \mathbf{P}$, we can define a generalized inverse $(\mathbf{S}_p \mathbf{V} \mathbf{S}_p)^{-g}$ such that

$$(\mathbf{S}_p \mathbf{V} \mathbf{S}_p)^{-g} = \mathbf{P}_p(\varrho * \text{diag } (\lambda_s) + \Sigma * \mathbf{I}_p)^{-1}\mathbf{P}_p .$$

Note that $(\mathbf{S}_p \mathbf{V} \mathbf{S}_p)^{-g}(\mathbf{S}_p \mathbf{V} \mathbf{S}_\nu) = \mathbf{S}_p$ and that multiplication by $\mathbf{S}_p$ does not change $(\mathbf{S}_p \mathbf{V} \mathbf{S}_p)^{-g}$.

The elements of $\mathbf{P}'_p \mathbf{y}$ are $p \times 1$ vectors $\mathbf{u}_s$ $(s = 1, \cdots, n - t)$ with variance-covariance matrices $\mathbf{V}_s = \lambda_s \varrho + \Sigma$. Hence the required log likelihood is given by

$$L' = \text{constant} - \tfrac{1}{2} \sum_s \log |\mathbf{V}_s| + \tfrac{1}{2}\mathbf{R}$$

where $\mathbf{R}$ is the weighted function of the $\mathbf{u}_s$ given by

$$\mathbf{R} = \sum_s \mathbf{u}'_s \mathbf{V}_s^{-1} \mathbf{u}_s = \mathbf{y}'(\mathbf{S}_p \mathbf{V} \mathbf{S}_p)^{-g}\mathbf{y}.$$

The estimates $\hat{\varrho}$, $\hat{\Sigma}$ maximizing L' are obtained by solving the equations

$$\frac{\partial L'}{\partial \rho_{ij}} = -E^*_{ij} + B^*_{ij} = 0 \qquad (i = 1, \cdots, p; j \leq i)$$

$$\frac{\partial L'}{\partial \Sigma_{ij}} = -F^*_{ij} + A^*_{ij} = 0 \qquad (i = 1, \cdots, p; j \leq i)$$

where

$$2E^*_{ij} = \text{tr } [(\mathbf{S}_p \mathbf{V} \mathbf{S}_p)^{-g} \mathbf{S}_p (\mathbf{C}_{ij} * \mathbf{ZZ'}) \mathbf{S}_p],$$

$$2B^*_{ij} = \mathbf{y}'(\mathbf{S}_p \mathbf{V} \mathbf{S}_p)^{-g} \mathbf{S}_p (\mathbf{C}_{ij} * \mathbf{ZZ'}) \mathbf{S}_p (\mathbf{S}_p \mathbf{V} \mathbf{S}_p)^{-g} \mathbf{y},$$

$$2F^*_{ij} = \text{tr } [(\mathbf{S}_p \mathbf{V} \mathbf{S}_p)^{-g} \mathbf{S}_p (\mathbf{C}_{ij} * \mathbf{I}_n) \mathbf{S}_p],$$

and

$$2A^*_{ij} = \mathbf{y}'(\mathbf{S}_p \mathbf{V} \mathbf{S}_p)^{-g} \mathbf{S}_p (\mathbf{C}_{ij} * \mathbf{I}_n) \mathbf{S}_p \mathbf{V} \mathbf{S}_p)^{-g} \mathbf{y}.$$

$\mathbf{D}_{ij}$ is a $p \times p$ matrix with one as the i, jth element and zero elsewhere and $\mathbf{C}_{ij} = \mathbf{D}_{ij} + \mathbf{D}_{ji} - \delta_{ij}\mathbf{D}_{ij}$ (where δ_{ij} is the Kronecker delta). (A_{ij}^* is related to the A_{ij}, defined in section 2, by the formula $A_{ij}^* = (2 - \delta_{ij})A_{ij}$, and similarly for **B**, **E**, and **F**).

Deviation of the results in section 4

In section 4, we discuss the application of the suggested method to parent-offspring data. We show here how the results quoted in section 4 can be derived.

We let the complete data vector **y**, of size $(N + m)p \times 1$, be given by

$$R_p[\mathbf{y}] = \mathbf{Y}$$

where **Y** is a $(N + m) \times p$ matrix with ith row

$$\mathbf{y}_{i1}(1 \leq i \leq n) \quad \text{and} \quad (i + N)\text{th row } \mathbf{y}_{i2}(1 \leq i \leq m).$$

The variance matrix of **y** can then be represented by **V**, where

$$\mathbf{V} = \mathbf{\Sigma} * \mathbf{I}_p + \varrho^* \mathbf{Z}\mathbf{Z}' \tag{A1}$$

and **Z** is a $(N + m) \times N$ matrix with one in the i, ith element ($i = 1, \cdots, N$) one in the $j + N, j$th element ($j = 1, \cdots, m$) and zero elsewhere. $\mathbf{Z}'\mathbf{Z}$ is then a $N \times N$ diagonal matrix with element (i, i) equal to two when $i \leq m$ and equal to one when $i > m$. We can write $\boldsymbol{\alpha}$ as

$$\boldsymbol{\alpha} = \boldsymbol{\mu}_1' * \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \boldsymbol{\mu}_2' * \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{A2}$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{1}_N & \mathbf{0}_N \\ \mathbf{0}_m & \mathbf{1}_m \end{bmatrix} \tag{A3}$$

so

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} N & 0 \\ 0 & m \end{bmatrix}$$

and

$$\mathbf{X}'\mathbf{Z} = \begin{bmatrix} \mathbf{1}_m' & \mathbf{1}_{N-m}' \\ \mathbf{1}_m' & \mathbf{0}_{N-m}' \end{bmatrix}$$

where $\mathbf{1}_q$ and $\mathbf{0}_q$ are $q \times 1$ vectors with respectively 1 and 0 in every element.

Eliminating $\hat{\boldsymbol{\beta}}$ from equations (11) we find $\hat{\boldsymbol{\alpha}}$ satisfies

$$\left[\mathbf{I}_p * \begin{pmatrix} N & 0 \\ 0 & m \end{pmatrix} - \hat{\mathbf{G}}_1 * \begin{pmatrix} N - m & 0 \\ 0 & 0 \end{pmatrix} - \hat{\mathbf{G}}_2 * \begin{pmatrix} m & m \\ m & m \end{pmatrix}\right]\boldsymbol{\alpha}$$

$$= N\bar{\mathbf{y}}_0' * \begin{pmatrix} 1 \\ 0 \end{pmatrix} + m\bar{\mathbf{y}}_2' * \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \hat{\mathbf{G}}_1[N\bar{\mathbf{y}}_0' - m\bar{\mathbf{y}}_1']$$

$$* \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \hat{\mathbf{G}}_2[m\bar{\mathbf{y}}_1' + m\bar{\mathbf{y}}_2'] * \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{A4}$$

where

$$\mathbf{G}_1 = (\mathbf{I}_p + \boldsymbol{\gamma}^{-1})^{-1} \quad \text{and} \quad \mathbf{G}_2 = (2\mathbf{I}_p + \boldsymbol{\gamma}^{-1})^{-1}.$$

Noting that the coefficients of $\hat{\boldsymbol{\alpha}}$ in equation (A4) are the elements of $((\mathbf{I}_p * \mathbf{X})\hat{\mathbf{H}}^{-1}(\mathbf{I}_p * \mathbf{X}'))$ and that the inverse of this is given by

$$((\mathbf{I}_p * \mathbf{X})\hat{\mathbf{H}}^{-1}(\mathbf{I}_p * \mathbf{X}'))^{-1}$$
$$= \mathbf{I}_p * \begin{pmatrix} 1/N & 0 \\ 0 & 1/m \end{pmatrix} + (1/N)\hat{\boldsymbol{\gamma}} * \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \left(\frac{N-m}{Nm}\right)\hat{\mathbf{G}}_1 * \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{(A5)}$$

we find

$$\hat{\boldsymbol{\alpha}} = \bar{\mathbf{y}}_0' * \begin{pmatrix} 1 \\ 0 \end{pmatrix} + [\bar{\mathbf{y}}_2' - \hat{\mathbf{G}}_1(\bar{\mathbf{y}}_1' - \bar{\mathbf{y}}_0')] * \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \text{(A6)}$$

and hence $\hat{\boldsymbol{\mu}}_1$ and $\hat{\boldsymbol{\mu}}_2$ are given by equations (22).

Using equations (11) to express $\boldsymbol{\beta}$ in terms of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ we find

$$\begin{aligned} 2\mathbf{B} &= (\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1}(\mathbf{s}^2 - \mathbf{M}_{11})(\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} \\ &+ (2\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1}(\mathbf{M}_{11} + \mathbf{M}_{12} + \mathbf{M}_{21} + \mathbf{M}_{22})(2\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} \\ &+ (\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1}[N(\bar{\mathbf{y}}_0 - \boldsymbol{\mu}_1)'(\bar{\mathbf{y}}_0 - \boldsymbol{\mu}_1) - m(\bar{\mathbf{y}}_1 - \boldsymbol{\mu}_1)'(\bar{\mathbf{y}}_1 - \boldsymbol{\mu}_1)](\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} \\ &+ (2\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1}[m(\bar{\mathbf{y}}_1 + \bar{\mathbf{y}}_2 - \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'(\bar{\mathbf{y}}_1 + \bar{\mathbf{y}}_2 - \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)](2\boldsymbol{\varrho} + \boldsymbol{\Sigma})^{-1} \end{aligned} \quad \text{(A7)}$$

and

$$\begin{aligned} 2\mathbf{A} &= \boldsymbol{\Sigma}^{-1}[\mathbf{s}^2 + \mathbf{M}_{22} - \mathbf{s}^2\mathbf{G}_1' + \mathbf{M}_{11}\mathbf{G}_1' - (\mathbf{M}_{11} + \mathbf{M}_{12} + \mathbf{M}_{21} + \mathbf{M}_{22})\mathbf{G}_2']\boldsymbol{\Sigma}^{-1} \\ &+ \boldsymbol{\Sigma}^{-1}[N(\bar{\mathbf{y}}_0)'(\bar{\mathbf{y}}_0 - \boldsymbol{\mu}_1)(\mathbf{I}_p - \mathbf{G}_1') + m(\bar{\mathbf{y}}_2)'(\bar{\mathbf{y}}_2 - \boldsymbol{\mu}_2) \\ &+ m(\bar{\mathbf{y}}_1)'(\bar{\mathbf{y}}_1 - \boldsymbol{\mu}_1)\mathbf{G}_1' - m(\bar{\mathbf{y}}_1 + \bar{\mathbf{y}}_2)'(\bar{\mathbf{y}}_1 + \bar{\mathbf{y}}_2 - \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\mathbf{G}_2']\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\varrho}\mathbf{B}. \end{aligned} \quad \text{(A8)}$$

Equations (A7) and (A8) simplify to equations (27) and (25) by using equations (A2) and (A6) and the identity

$$(2\mathbf{I}_p + \boldsymbol{\gamma}^{-1}) = \tfrac{1}{2}\mathbf{I}_p - \tfrac{1}{2}(2\boldsymbol{\gamma} + \mathbf{I}_p)^{-1}.$$

Substituting (A5) in equation (22) we find $\mathbf{E}$ and $\mathbf{F}$ are given by equations (28) and (26).

Derivation of the results in section 5

Section 5 discusses the case when there is selection of parents on the basis of their records. We first show that the likelihood in this case is equivalent to that used in section 4.

Using the vectors defined in section 4 we know the N vectors $\mathbf{y}_{i1}$ ($i = 1, \cdots, N$) are normally distributed with mean $\boldsymbol{\mu}_1$ and variance matrix $\boldsymbol{\Sigma} + \boldsymbol{\varrho}$. Because selection is based on $\mathbf{y}_{i1}$ the distribution of $\mathbf{y}_{i2}$ is independent of the first record with mean $\boldsymbol{\mu}_2 + (\mathbf{y}_{i1} - \boldsymbol{\mu}_1)(\boldsymbol{\Sigma} + \boldsymbol{\varrho})^{-1}\boldsymbol{\varrho}$ and variance matrix $\boldsymbol{\Sigma} + \boldsymbol{\varrho} - \boldsymbol{\varrho}(\boldsymbol{\Sigma} + \boldsymbol{\varrho})^{-1}\boldsymbol{\varrho}$.

The log likelihood L^* can then be written as

$$L^* = \text{constant} - \frac{N}{2}\log|\boldsymbol{\Sigma}+\boldsymbol{\varrho}| + \sum_{i=1}^{N}(\mathbf{y}_{i1}-\boldsymbol{\mu}_1)(\boldsymbol{\Sigma}+\boldsymbol{\varrho})^{-1}(\mathbf{y}_{i1}-\boldsymbol{\mu}_1)'$$

$$-\frac{m}{2}\log|\boldsymbol{\Sigma}+\boldsymbol{\varrho}-\boldsymbol{\varrho}(\boldsymbol{\Sigma}+\boldsymbol{\varrho})^{-1}\boldsymbol{\varrho}|$$

$$+\sum_{i=1}^{m}(\mathbf{y}_{i2}-\boldsymbol{\mu}_2-(\mathbf{y}_{i1}-\boldsymbol{\mu}_1)(\boldsymbol{\Sigma}+\boldsymbol{\varrho})^{-1}\boldsymbol{\varrho})(\boldsymbol{\Sigma}+\boldsymbol{\varrho}-\boldsymbol{\varrho}(\boldsymbol{\Sigma}+\boldsymbol{\varrho})^{-1}\boldsymbol{\varrho})^{-1}$$

$$\cdot(\mathbf{y}_{i2}-\boldsymbol{\mu}_2-(\mathbf{y}_{i1}-\boldsymbol{\mu}_1)(\boldsymbol{\Sigma}+\boldsymbol{\varrho})^{-1}\boldsymbol{\varrho})'.$$

This is a factorization of the log likelihood (5) when $\mathbf{V}$, $\mathbf{X}$, and $\boldsymbol{\alpha}$ are as defined in the appendix (Equations (A1)–(A3)). Hence equations (21) to (26) can be used as a basis to estimate $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\boldsymbol{\varrho}$, and $\boldsymbol{\Sigma}$ even if the parents that have offspring recorded are selected.

However the information matrix depends on the expectation of the moments $\mathbf{M}_{11}$, $\mathbf{M}_{12}$, $\mathbf{M}_{21}$, and $\mathbf{M}_{22}$, and these depend on the selection scheme and we will outline the modifications.

First we find the expected values of $\mathbf{s}^2$, $\mathbf{M}_{12}$, $\mathbf{M}_{21}$, and $\mathbf{M}_{22}$. $\mathbf{M}_{11}$ depends on the selection scheme and will be taken as fixed. $\mathbf{s}^2$ is independent of the selection scheme and so $E(\mathbf{s}^2) = (N-1)(\boldsymbol{\Sigma}+\boldsymbol{\varrho})$. Since $\mathbf{y}_{i2}$ can be written as

$$\mathbf{y}_{i2} = \boldsymbol{\mu}_2 - [\mathbf{y}_{i1}-\boldsymbol{\mu}_1](\boldsymbol{\Sigma}+\boldsymbol{\varrho})^{-1}\boldsymbol{\varrho} + \boldsymbol{\varepsilon}_{i2.1}$$

where $\boldsymbol{\varepsilon}_{i2.1}$ is normally distributed with variance matrix $(\boldsymbol{\Sigma}+\boldsymbol{\varrho}) - \boldsymbol{\varrho}(\boldsymbol{\Sigma}+\boldsymbol{\varrho})^{-1}\boldsymbol{\varrho}$.

So $E(\mathbf{M}_{22}) = \boldsymbol{\varrho}(\boldsymbol{\Sigma}+\boldsymbol{\varrho})^{-1}\mathbf{M}_{11}(\boldsymbol{\Sigma}+\boldsymbol{\varrho})^{-1}\boldsymbol{\varrho} + (n-1)((\boldsymbol{\Sigma}+\boldsymbol{\varrho}) - \boldsymbol{\varrho}(\boldsymbol{\Sigma}+\boldsymbol{\varrho})^{-1}\boldsymbol{\varrho})$ (A9)

$$E(\mathbf{M}_{12}) = \mathbf{M}_{11}(\boldsymbol{\Sigma}+\boldsymbol{\varrho})^{-1}\boldsymbol{\varrho} \tag{A10}$$

and

$$E(\mathbf{M}_{21}) = \boldsymbol{\varrho}(\boldsymbol{\Sigma}+\boldsymbol{\varrho})^{-1}\mathbf{M}_{11}\,. \tag{A11}$$

It is reassuring to find that the expected values of $\mathbf{A}$ and $\mathbf{B}$ are $\mathbf{F}$ and $\mathbf{E}$ (equations (26) and (28)) even when there is selection.

Now differentiating $2\mathbf{A}$ with respect to $\boldsymbol{\Sigma}_{ij}$ $(i = 1, \cdots, p; j \leq i)$ we find

$$2\frac{\partial \mathbf{A}}{\partial \boldsymbol{\Sigma}_{ij}} = \tfrac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbf{C}_{ij}\boldsymbol{\Sigma}^{-1}(\mathbf{M}_{11}-\mathbf{M}_{12}-\mathbf{M}_{21}-\mathbf{M}_{22})\boldsymbol{\Sigma}^{-1}$$

$$-\tfrac{1}{2}\boldsymbol{\Sigma}^{-1}(\mathbf{M}_{11}-\mathbf{M}_{12}-\mathbf{M}_{21}+\mathbf{M}_{22})\boldsymbol{\Sigma}^{-1}\mathbf{C}_{ij}\boldsymbol{\Sigma}^{-1}$$

$$-(\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}\mathbf{C}_{ij}(\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}(\mathbf{s}^2-\mathbf{M}_{11})(\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}$$

$$-(\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}(\mathbf{s}^2-\mathbf{M}_{11})(\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}\mathbf{C}_{ij}(\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}$$

$$-\tfrac{1}{2}(2\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}\mathbf{C}_{ij}(2\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}(\mathbf{M}_{11}+\mathbf{M}_{12}+\mathbf{M}_{21}+\mathbf{M}_{22})(2\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}$$

$$-\tfrac{1}{2}(2\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}(\mathbf{M}_{11}+\mathbf{M}_{12}+\mathbf{M}_{21}+\mathbf{M}_{22})(2\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}\mathbf{C}_{ij}(2\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}.$$

Using (A9) to (A11) we find this has expected value given by

$$2E\left(\frac{\partial \mathbf{A}}{\partial \boldsymbol{\Sigma}_{ij}}\right) = -2(m-1)\boldsymbol{\Sigma}^{-1}\mathbf{C}_{ij}\boldsymbol{\Sigma}^{-1} - 2(N-m)(\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}\mathbf{C}_{ij}(\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}$$
$$- 2(m-1)(2\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}\mathbf{C}_{ij}(2\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}$$
$$- \tfrac{1}{2}[\boldsymbol{\Sigma}^{-1} + (2\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1} - 2(\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}]\mathbf{C}_{ij}\mathbf{M}_{11}^{*}$$
$$- \tfrac{1}{2}\mathbf{M}_{11}^{*}\mathbf{C}_{ij}[\boldsymbol{\Sigma}^{-1} + (2\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1} - 2(\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}].$$

Hence $\mathbf{J}_{22}^{*} = \mathbf{J}_{22} + [\boldsymbol{\Sigma}^{-1} + (2\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1} - 2(\boldsymbol{\varrho}+\boldsymbol{\Sigma})^{-1}]^{*}\mathbf{M}_{11}^{*}$. Similarly by differentiating $\mathbf{B}$ with respect to $\boldsymbol{\Sigma}_{ij}$ and ρ_{ij} we find $\mathbf{J}_{11}^{*}$ and $\mathbf{J}_{21}^{*}$ as given in equation (30).

Received January 1972, *Revised July* 1972

Key Words: Multivariate variance components; Records subject to culling; Maximum likelihood.

Reprinted from
BIOMETRICS Copyright © 1973
THE BIOMETRIC SOCIETY, Vol. 29, No. 3, September 1973

6

Simultaneous estimation of variance and covariance components from multitrait mixed model equations

by

L.R. Schaeffer, J.W. Wilton and R. Thompson

Reprinted from
BIOMETRICS COPYRIGHT © 1978
THE BIOMETRIC SOCIETY, Vol. 34, No. 2, June 1978

BIOMETRICS 34, 199–208
June, 1978

Simultaneous Estimation of Variance and Covariance Components from Multitrait Mixed Model Equations

L. R. SCHAEFFER and J. W. WILTON

Department of Animal and Poultry Science, University of Guelph
Guelph, Ontario, Canada

R. THOMPSON

ARC Unit of Statistics, University of Edinburgh
Mayfield Road, Edinburgh EH9 3JZ, Scotland

Summary

A procedure is described for estimating variance and covariance components when different variables are observed on different experimental units. The procedure allows for different linear models in the different traits. The method consists of maximizing the likelihood of a set of error contrasts. It is an extension of a procedure presented by Thompson (1973) for the case when all variables are measured on all traits. The method is also an iterative version of Rao's (1971) minimum norm quadratic unbiased estimation procedure (MINQUE). The calculations are described in terms of solutions to Henderson's (1973) mixed model equations. The procedure was developed to estimate sire components of variance and covariance for yearling weights of male and female progeny in beef cattle, and a sample of such data is used in a worked example.

1. Introduction

Components of covariance are usually estimated between traits measured on the same individual or experimental unit when the same linear model is assumed for each trait. For example, the genetic covariance between milk and fat production in dairy cattle is obtained from measurements of both milk and fat yield on each cow. This type of covariance is estimated using either the sum or the crossproduct of the two traits of interest (Kempthorne, pp. 264–267, 1969).

Sometimes, however, crossproducts or sums of two traits do not exist. Yearling weight, for example, is measured on male and female beef calves. The male and female measurements are usually analyzed separately because the sexes are raised under differing environmental conditions (i.e. different feedlots or pastures), and because the variability of yearling weight is usually greater for males than females. Thus, beef sires can be evaluated twice, from their male progeny and their female progeny, for their estimated transmitting ability for yearling weight. In this case each experimental unit (i.e. progeny of a sire) has measurements on one trait and the usual estimation procedure for covariance components, i.e. between male and female yearling weight, using crossproducts or sums on each unit, is not possible.

This paper presents a method for simultaneously estimating components of variance and covariance when crossproducts or sums between two traits do not exist. Perhaps the simplest

Key Words: Variance and covariance components; Simultaneous estimation; Multiple traits; Modified maximum likelihood.

explanation of the method is to say that one first solves Henderson's (1973) mixed model equations, and then equates quadratic forms of these solutions to their expectation. This is how the equations in section 3 are presented. We show the method to be an extension of a maximum likelihood (ML) procedure derived by Thompson (1973), who considered the case when the same model was appropriate to all traits and all traits were measured on all individuals. We also show that the method can be related to the minimum norm quadratic unbiased estimation (MINQUE) method of Rao (1971). In Section 4 we give an example of estimating the sire components of variance and covariance for male and female yearling weights in beef cattle.

2. *Biological Situation*

Exotic beef cattle from Europe have made a radical impact on the Canadian and United States' beef populations and markets. Many European bulls were imported to North America without any accompanying growth information. Progeny testing programs were organized to evaluate a sire's genetic potential for growth as well as other traits not directly measurable on the sire itself. Yearling weight is one of the growth traits of interest to cattlemen. Males are slaughtered at or shortly after yearling age while females are kept in the herd to produce future calves. Cattlemen would prefer to have fast growing male calves and slower growing female calves. If such an interaction of sire and sex of calf were present it would be indicated by a sire component of covariance that yielded a genetic correlation less than one.

Table 1 shows a summary of sire by environment subclass totals for male and female yearling weights of progeny of four sires made in four different environments. These records are a sample of Charolais beef data from the Canadian Record of Performance program. The definition of an environment includes herd, year of calving, and feeding system interaction and is not critical to the discussion of methods of estimation. It is important to note, however, that the environmental effects are not the same for both male and female progeny. Male and female calves are commonly separated at weaning and provided with different feeding and management conditions. The total sums of squares for male and female yearling weights are also shown in Table 1, and these figures will be needed later.

3. *Theoretical Development*

3.1. *Model and Assumptions*

An appropriate model might be

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e} \tag{1}$$

where $\mathbf{y}$, the $(N \times 1)$ vector of observations, can be written in terms of t $(N_i \times 1)$ vectors, $\mathbf{y}_i$, representing observations on the t traits. Thus $\mathbf{y}' = (\mathbf{y}_1', \cdots, \mathbf{y}_t')$. The matrices $\mathbf{X}$ and $\mathbf{Z}$ can be written as

$$\mathbf{X} = \sum_{i=1}^{t}{}^{+} \mathbf{X}_i \quad \text{and} \quad \mathbf{Z} = \sum_{i=1}^{t}{}^{+} \mathbf{Z}_i,$$

where $\sum^{+}$ represents the direct sum (Searle, p. 213, 1966). Without loss of generality each matrix $\mathbf{X}_i$ (of size $N_i \times p_i$) is assumed to be of full column rank. The vectors of fixed and random effects, $\mathbf{b}$ and $\mathbf{u}$, can be written as $\mathbf{b}' = (\mathbf{b}_1', \cdots, \mathbf{b}_t')$ and $\mathbf{u}' = (\mathbf{u}_1', \cdots, \mathbf{u}_t')$. The

TABLE 1

Sire by Environment Subclass Totals for Male and Female Progeny Yearling Weights (kg).

Environments	Sires 1	2	3	4
Males				
1	3994 (9)[a]	996 (2)	2073 (4)	491 (1)
2	945 (2)	828 (2)	893 (2)	1309 (3)
3	782 (2)	694 (2)	2157 (5)	1722 (4)
4	846 (2)	409 (1)	863 (2)	0 (0)
Females				
5	4080 (10)	1088 (3)	1963 (5)	371 (1)
6	1971 (5)	1711 (5)	1485 (4)	735 (2)
7	625 (2)	875 (3)	1561 (4)	1014 (3)
8	685 (2)	307 (1)	1951 (6)	0 (0)
Total sums of squares (kg^2)				
Males	8,546,828			
Females	7,580,124			

[a] number of observations in parentheses

length of $\mathbf{b}_i$ is p_i. Each random effect vector, $\mathbf{u}_i$, contains only one random factor and each has the same number of levels, say s. There is assumed to be at least one record of the random factor for at least one trait. Finally, $\mathbf{e}' = (\mathbf{e}_1', \cdots, \mathbf{e}_t')$.

We assume that $E(\mathbf{y}_i) = \mathbf{X}_i\mathbf{b}_i$, $E(\mathbf{u}_i) = \mathbf{0}$ and $E(\mathbf{e}_i) = \mathbf{0}$ $(i = 1, 2 \cdots t)$. We also assume that the variance matrices of $(\mathbf{u}_i, \mathbf{u}_j)$, $(\mathbf{e}_i, \mathbf{e}_j)$ (for $i \geq j = 1, 2 \cdots t$) and $(\mathbf{u}, \mathbf{e})$ are

$$\begin{bmatrix} \mathbf{I}\sigma_{ii} & \mathbf{I}\sigma_{ij} \\ \mathbf{I}\sigma_{ij} & \mathbf{I}\sigma_{jj} \end{bmatrix}, \begin{bmatrix} \mathbf{I}\sigma_{ei}^2 & \mathbf{O} \\ \mathbf{O} & \mathbf{I}\sigma_{ej}^2 \end{bmatrix} \text{ and } \begin{bmatrix} \mathbf{G} & \mathbf{O} \\ \mathbf{O} & \mathbf{R} \end{bmatrix} \tag{2}$$

We will find it useful to write the $t(t + 3)/2$ parameters indexed by a single parameter and so we let $\boldsymbol{\theta}$ be a $(t(t + 3)/2 \times 1)$ vector with elements $\theta_{i(i-1)/2+j} = \sigma_{ij}$, $\theta_{t(t+1)/2+i} = \sigma_{ei}^2$ $(i \geq j = 1, 2 \cdots t)$. For example, when $t = 2$, $\boldsymbol{\theta}' = (\sigma_{11}, \sigma_{21}, \sigma_{22}, \sigma_{e1}^2, \sigma_{e2}^2)$. We let $\mathfrak{G}$ be a symmetric matrix with $(i, j)th$ and $(j, i)th$ elements equal to $\sigma_{ij}(i \geq j)$ and let $\mathfrak{G}_e^2$ be a $(t \times t)$ diagonal matrix with ith diagonal element σ_{ei}^2. Hence, the rows of the lower triangular part of $\mathfrak{G}$ appear in order as the first $t(t + 1)/2$ elements of $\boldsymbol{\theta}$.

The variance-covariance matrix of $\mathbf{y}$, $\mathbf{V}$ can now be written as a linear function of these unknown parameters, i.e.,

$$\mathbf{V} = \sum_{i \geq j=1}^{t} \mathbf{Z}(\mathbf{I}*\mathbf{D}_{ij})\mathbf{Z}'\sigma_{ij} + \sum_{i=1}^{t} \mathbf{D}_i\sigma_{ei}^2 = \sum_{k=1}^{t(t+3)/2} \mathbf{V}_k\theta_k. \tag{3}$$

We use $*$ to represent the direct product (Searle, p. 215, 1966) and $\mathbf{D}_{ij}$ is a $t \times t$ matrix with $(i, j)th$ and $(j, i)th$ elements equal to one and all others zero. $\mathbf{I}$ is a $s \times s$ identity matrix and $\mathbf{D}_i$ is an $N \times N$ diagonal matrix with the jth diagonal element equal to one if the jth observation is on the ith trait and zero otherwise.

The calculations will be described in terms of this specific model. Some immediate, but more complex, extensions of this model would allow more than one random factor in each $\mathbf{u}_i$.

3.2. Mixed Model Equations

The procedure we describe is iterative and must be started by assigning initial values to the unknown components. We use boldface to represent functions of these initial values, i.e. $\tilde{\boldsymbol{\sigma}}$ and $\tilde{\boldsymbol{\sigma}}_e^2$ are initial estimates of $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}_e^2$. We let

$$\mathbf{B} = \tilde{\mathbf{G}}^{-1} = \begin{bmatrix} \mathbf{I}\tilde{\sigma}_{11} & \mathbf{I}\tilde{\sigma}_{21} & \cdots & \mathbf{I}\tilde{\sigma}_{t1} \\ \mathbf{I}\tilde{\sigma}_{21} & \mathbf{I}\tilde{\sigma}_{22} & \cdots & \mathbf{I}\tilde{\sigma}_{t2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \mathbf{I}\tilde{\sigma}_{t1} & \mathbf{I}\tilde{\sigma}_{t2} & \cdots & \mathbf{I}\tilde{\sigma}_{tt} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I}\alpha_{11} & \mathbf{I}\alpha_{21} & \cdots & \mathbf{I}\alpha_{t1} \\ \mathbf{I}\alpha_{21} & \mathbf{I}\alpha_{22} & \cdots & \mathbf{I}\alpha_{t2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \mathbf{I}\alpha_{t1} & \mathbf{I}\alpha_{t2} & \cdots & \mathbf{I}\alpha_{tt} \end{bmatrix}, \quad (4)$$

and

$$\mathbf{R}^{-1} = \begin{bmatrix} \mathbf{I}\tilde{\sigma}_{e1}^2 & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{I}\tilde{\sigma}_{e2}^2 & \cdots & \mathbf{O} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \mathbf{O} & \mathbf{O} & \cdots & \mathbf{I}\tilde{\sigma}_{et}^2 \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I}\gamma_1 & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{I}\gamma_2 & \cdots & \mathbf{O} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \mathbf{O} & \mathbf{O} & \cdots & \mathbf{I}\gamma_t \end{bmatrix}, \quad (5)$$

We note that $\mathbf{B}$ is the inverse of the direct product of $\tilde{\boldsymbol{\sigma}}$ times an identity matrix, i.e. $\mathbf{B} = \mathbf{I} * \tilde{\boldsymbol{\sigma}}^{-1}$. Thus, $\mathbf{B}$ only exists if $\tilde{\boldsymbol{\sigma}}$ is non-singular. We will assume for the moment that $\tilde{\boldsymbol{\sigma}}$ is non-singular, but will later relax this assumption. We note that $\tilde{\boldsymbol{\sigma}}$ is of size $t \times t$ and t will usually be small in relation to $\mathbf{B}$.

The initial values for the variances may be obtained from prior analyses where each trait was considered individually. Previous experience might suggest initial values for the covariances, and obviously, the closer the initial values are to the final estimates, the fewer rounds of iteration needed.

We present the calculations in terms of the mixed model equations of Henderson (1973). These are of the form

$$\begin{bmatrix} \mathbf{X}'\tilde{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\tilde{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\tilde{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\tilde{\mathbf{R}}^{-1}\mathbf{Z} + \tilde{\mathbf{G}}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\tilde{\mathbf{R}}^{-1}\mathbf{y} \\ \mathbf{Z}'\tilde{\mathbf{R}}^{-1}\mathbf{y} \end{bmatrix}. \quad (6)$$

These equations are a convenient way of solving the generalized least squares equations

$$\hat{\mathbf{b}} = (\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{y}. \quad (7)$$

The term $\hat{\mathbf{u}}$ can be thought of as a predictor of the random effects vector $\hat{\mathbf{u}}$. Using the fact that $\mathbf{X}'\mathbf{X}$ and $\tilde{\mathbf{R}}$ are block diagonal, equation (6) can be set out and is shown below in the form for two traits:

$$\begin{bmatrix} \gamma_1\mathbf{X}_1'\mathbf{X}_1 & \mathbf{O} & \gamma_1\mathbf{X}_1'\mathbf{Z}_1 & \mathbf{O} \\ \mathbf{O} & \gamma_2\mathbf{X}_2'\mathbf{X}_2 & \mathbf{O} & \gamma_2\mathbf{X}_2'\mathbf{Z}_2 \\ \gamma_1\mathbf{Z}_1'\mathbf{X}_1 & \mathbf{O} & \gamma_1\mathbf{Z}_1'\mathbf{Z}_1 + \mathbf{I}\alpha_{11} & \mathbf{I}\alpha_{21} \\ \mathbf{O} & \gamma_2\mathbf{Z}_2'\mathbf{X}_2 & \mathbf{I}\alpha_{21} & \gamma_2\mathbf{Z}_2'\mathbf{Z}_2 + \mathbf{I}\alpha_{22} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \\ \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \end{bmatrix} = \begin{bmatrix} \gamma_1\mathbf{X}_1'\mathbf{y}_1 \\ \gamma_2\mathbf{X}_1'\mathbf{y}_2 \\ \gamma_1\mathbf{Z}_1'\mathbf{y}_1 \\ \gamma_2\mathbf{Z}_2'\mathbf{y}_2 \end{bmatrix}. \quad (8)$$

The number of levels in $\mathbf{b}_i$ may sometimes be large, and hence, the equations may be difficult to solve directly. If this is a problem a solution is to absorb the equations of the fixed effects, $\hat{\mathbf{b}}_i$, into the $\hat{\mathbf{u}}_i$ equations. This gives

$$\begin{bmatrix} \mathbf{Q}_1 + \mathbf{I}\alpha_{11} & \mathbf{I}\alpha_{21} & \cdots & \mathbf{I}\alpha_{t1} \\ \mathbf{I}\alpha_{21} & \mathbf{Q}_2 + \mathbf{I}\alpha_{22} & \cdots & \mathbf{I}\alpha_{t2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \mathbf{I}\alpha_{t1} & \mathbf{I}\alpha_{t2} & \cdots & \mathbf{Q}_t + \mathbf{I}\alpha_{tt} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\mathbf{u}}_t \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{r}_t \end{bmatrix}, \tag{9}$$

where $\mathbf{Q}_i = \gamma_i \mathbf{Z}_i' \mathbf{P}_i \mathbf{Z}_i, \mathbf{r}_i = \gamma_i \mathbf{Z}_i' \mathbf{P}_i \mathbf{y}_i$ and $\mathbf{P}_i = \mathbf{I} - \mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'$. If we let

$$\mathbf{Q} = \sum_{i=1}^{t}{}^{+} \mathbf{Q}_i \quad \text{and} \quad \mathbf{P} = \sum_{i=1}^{t}{}^{+} \mathbf{P}_i$$

(9) can be written as $(\mathbf{Q} + \mathbf{B})\hat{\mathbf{u}} = \mathbf{r}$, where $\mathbf{r} = \hat{\mathbf{R}}^{-1}\mathbf{Z}'\mathbf{P}\mathbf{y}$. The inverse of the coefficient matrix in (9), $(\mathbf{Q} + \mathbf{B})$, can be represented by the symmetric matrix

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1t} \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \cdots & \mathbf{C}_{2t} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \mathbf{C}_{t1} & \mathbf{C}_{t2} & \cdots & \mathbf{C}_{tt} \end{bmatrix}. \tag{10}$$

$\mathbf{C}$ is equivalent to the segment of the inverse of the coefficient matrix in (8) corresponding to the $\hat{\mathbf{u}}$ rows and columns.

3.3 Variance and Covariance Component Estimation

In this section we suggest an iterative scheme for estimating the variance parameters. The procedure is equivalent to estimating $\boldsymbol{\theta}$ by equating to their expectation the quadratic forms

$$\mathbf{u}_i'\mathbf{u}_j \text{ and } \mathbf{e}_i'\mathbf{e}_i \ (i \geq j = 1, \cdots, t), \tag{11}$$

regarding $\tilde{\boldsymbol{\sigma}}$ and $\tilde{\boldsymbol{\sigma}}_e{}^2$ as fixed quantities not necessarily equal to the underlying true values where $\hat{\mathbf{e}}_i$, which equals $\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}_i - \mathbf{Z}_i\hat{\mathbf{u}}_i$, is a 'residual' vector for the *ith* trait.

The calculations are then repeated using the new estimate of $\boldsymbol{\theta}$ in place of $\boldsymbol{\theta}$ until the estimates converge. Of course equating quadratic forms (11) to their expectation is equivalent to equating linear functions of (11) to their expectation. We find it more convenient to work with linear functions of (11) found by letting $\hat{\mathbf{v}}_i = \boldsymbol{\Sigma}_j \alpha_{ij}\hat{\mathbf{u}}_j$ and using the quadratic forms

$$(1 - \delta_{ij}/2)\hat{\mathbf{v}}_i'\hat{\mathbf{v}}_j = f_{i(i-1)/2+j}$$

and

$$(1/2)\gamma_i{}^2\hat{\mathbf{e}}_i'\hat{\mathbf{e}}_i = f_{t(t+1)/2+i} \ (i \geq j = 1, \cdots, t), \tag{12}$$

where δ_{ij} is the Kronecker delta.

Our first arguments for using these terms were rather intuitive, in that terms like $\hat{\mathbf{u}}_i$ are already used by animal breeders to predict breeding values when the variance parameters are known, and this prompted us to consider using functions of $\hat{\mathbf{u}}_i$ to estimate the variance parameters when the parameters are unknown.

There is some stronger justification for the method if we assume the variables are normally distributed. Patterson and Thompson (1971) in a similar situation suggested

maximizing the likelihood of error contrasts, i.e., contrasts of zero expectation, to estimate variance parameters. This has been called a restricted maximum likelihood approach (REML) in the review paper of Harville (1977). We will show that our iterative procedure converges (if it converges) to a solution of these REML estimating equations.

We note that $\mathbf{V}$ is linear in the parameters $\boldsymbol{\theta}$ and using arguments similar to Anderson (1973) it can be shown (Patterson and Thompson 1974 and Harville 1977) that the REML estimators satisfy equations of the form

$$\mathbf{F}\boldsymbol{\theta} = \mathbf{f} \tag{13}$$

where $\mathbf{F}$ is the information matrix of $\boldsymbol{\theta}$ and $\mathbf{f}$ is a vector of quadratic forms. Also if $\mathbf{F}^{(k)}$ and $\mathbf{f}^{(k)}$ are $\mathbf{F}$ and $\mathbf{f}$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$ then $\mathbf{E}(\mathbf{f}^{(k)}) = \mathbf{F}^{(k)}\boldsymbol{\theta}$. As (13) usually cannot be solved explicitly, an iterative scheme is necessary. The obvious scheme of defining the $(k+1)th$ iterate to be $\boldsymbol{\theta}^{(k+1)} = (\mathbf{F}^{(k)})^{-1}\mathbf{f}^{(k)}$, i.e. equating $f^{(k)}$ to its expectation, is equivalent to using Fisher's method of scoring (see Harville 1977).

For our particular application it is not very hard to verify that the elements of $\mathbf{f}$ are in fact the quadratic forms f_i given in (12) (see Section 5 in Harville 1977, noting that his $\bar{\mathbf{v}}$ defined in (3.2) is equivalent to a vector formed from our $\mathbf{v}_i$). Hence our method can be thought of as a REML procedure and corresponds to using the method of scoring to find solutions to the estimating equations.

We note also that the quadratic forms we use arise in Rao's minimum norm quadratic unbiased estimation (MINQUE) procedure (Harville 1977) and this gives a different justification for using the method.

We now give convenient computing forms for the terms in (13). We first note that a simplified computing algorithm for $\gamma_i^2\hat{\mathbf{e}}_i\hat{\mathbf{e}}_i$ is $\gamma_i^2\mathbf{y}_i'\mathbf{P}_i\mathbf{y}_i - \gamma_i^2\hat{\mathbf{u}}_i'\mathbf{Z}_i'\mathbf{P}_i'\mathbf{y}_i - \boldsymbol{\Sigma}_j(\alpha_{ij}\gamma_i)\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_j$ (see (A1) in appendix). We define three $(st \times st)$ matrices.

$$\mathbf{U} = \mathbf{B} - \mathbf{BCB},\ \mathbf{R}_\bullet^{-1} = (\mathbf{I} * \tilde{\boldsymbol{\delta}}_e^2)^{-1} \text{ and } \mathbf{T} = \mathbf{BCR}_\bullet^{-1}. \tag{14}$$

where $\mathbf{U}_{ij} = \mathbf{B}_{ij} - \boldsymbol{\Sigma}_k\boldsymbol{\Sigma}_l\alpha_{ik}\alpha_{jl}\mathbf{C}_{kl}$ and

$$\mathbf{T}_{ij} = \sum_{k=1}^{t} \alpha_{ik}\,\gamma_j\,\mathbf{C}_{kj}.$$

Then, as we show in the appendix ((A2)–(A4)), the expected values of the quadratic forms in (12) can be written using

$$E(\hat{\mathbf{v}}_i'\hat{\mathbf{v}}_i) = \sum_{k=1}^{t} tr(\mathbf{U}_{ki}\mathbf{U}_{ik})\sigma_{kk} + \sum_{l>k=1}^{t} 2tr(\mathbf{U}_{ki}\mathbf{U}_{il})\sigma_{lk} + \sum_{k=1}^{t} tr(\mathbf{U}_{ki}\mathbf{T}_{ik})\sigma_{ek}^2, \tag{15}$$

$$E(2\hat{\mathbf{v}}_i'\hat{\mathbf{v}}_j) = \sum_{k=1}^{t} 2tr(\mathbf{U}_{ki}\mathbf{U}_{jk})\sigma_{kk} + \sum_{l>k=1}^{t} 2tr(\mathbf{U}_{ki}\mathbf{U}_{jl} + \mathbf{U}_{kj}\mathbf{U}_{il})\sigma_{lk} + \sum_{k=1}^{t} tr(\mathbf{U}_{ki}\mathbf{T}_{jk} + \mathbf{U}_{kj}\mathbf{T}_{ik})\sigma_{ek}^2, \tag{16}$$

and

$$E(\gamma_i^2\hat{\mathbf{e}}_i'\hat{\mathbf{e}}_i) = \sum_{k=1}^{t} tr(\mathbf{T}_{ki}\mathbf{U}_{ik})\sigma_{kk} + \sum_{l>k=1}^{t} tr(\mathbf{T}_{ki}\mathbf{U}_{il} + \mathbf{T}_{li}\mathbf{U}_{ik})\sigma_{lk} + (N_i - P_i - s)\gamma_i^2\sigma_{ei}^2 + \sum_{k=1}^{t} tr(\mathbf{T}_{ki}\mathbf{T}_{ik})\sigma_{ek}^2. \tag{17}$$

Hence, we can find F_{ij} the coefficient of the parameters θ_j in the expected value of the quadratic forms f_i. For example, when $t = 2$, $2\mathbf{F}$ is given by

$$\begin{bmatrix} \mathrm{tr}(U_{11}U_{11}) & 2\mathrm{tr}(U_{11}U_{12}) & \mathrm{tr}(U_{21}U_{12}) & \mathrm{tr}(U_{11}T_{11}) & \mathrm{tr}(U_{21}T_{12}) \\ & 2\mathrm{tr}(U_{11}U_{22}+U_{12}U_{21}) & 2\mathrm{tr}(U_{12}U_{22}) & \mathrm{tr}(U_{11}T_{21}+U_{12}T_{11}) & \mathrm{tr}(U_{21}T_{22}+U_{22}T_{12}) \\ & & \mathrm{tr}(U_{22}U_{22}) & \mathrm{tr}(U_{12}T_{21}) & \mathrm{tr}(U_{22}T_{22}) \\ & & & \mathrm{tr}(T_{11}T_{11})+ & \mathrm{tr}(T_{21}T_{12}) \\ & \text{symmetric} & & (N_1-p_1-s)\gamma_1^2 & \\ & & & & \mathrm{tr}(T_{22}T_{22})+ \\ & & & & (N_2-p_2-s)\gamma_2^2 \end{bmatrix}$$

The matrices **U** and **F** are symmetric, and hence, not all the terms in (13) need to be calculated.

The asymptotic variance-covariance matrix, the inverse of the information matrix, can be calculated as $\mathbf{F}^{-1}$. In some circumstances **F** can be singular and this means that not all the parameters can be estimated. For instance, if the measurements of traits i and j have no random levels in common then σ_{ij} cannot be estimated.

3.4. Discussion

The computational requirements of one round of iteration of the proposed method and Henderson's (1953) Method 3 are of the same order of magnitude. Recall that Method 3 estimation requires solutions to equations similar to (9) with the elements of **B** put equal to zero. Then quadratic forms, functions of these solutions and **Q**, would be equated to their expectation.

The assumption that $\tilde{\mathfrak{d}}$ was nonsingular was introduced to enable the presentation of the results in terms of Henderson's mixed model equations. In fact we need make no assumption about the singularity of $\tilde{\mathfrak{d}}$ provided that we calculate **T** as $(\mathbf{Z'PZ}\tilde{\mathbf{G}}+\tilde{\mathbf{R}}_*)^{-1}$, $\hat{\mathbf{v}}$ as $\mathbf{TZ'Py}$, $\hat{\mathbf{u}}$ as $\tilde{\mathbf{G}}\tilde{\mathbf{v}}$ and **U** as $\mathbf{Z'PZT'}$ (see Thompson 1973 and Harville 1977).

Thompson (1973) considered the case when the same model was appropriate to all traits and allowed for the estimation of the covariance between $\mathbf{e}_i$ and $\mathbf{e}_j$. If we assume that this covariance is zero in Thompson's results, and let $\mathbf{X}_i=\mathbf{X}_1$, $\mathbf{Z}_i=\mathbf{Z}_1$, $N_i=N_1$ and $P_i=P_1$ in our results, then we find that the two methods are equivalent. Thompson's p, n, t, b, ρ_{ij}, $\mathbf{\Sigma}_{ii}$, **X**, **Z**, **S**, **U**, $(\varrho^{-1}*\mathbf{I}_b)\mathbf{W}^{-1}$, α, β and β^* are equivalent to t, N_1, p_1, s, σ_{ij}, σ^2_{ei}, $\mathbf{X}_1$, $\mathbf{Z}_1$, p, U, **T**, **b**, **u** and **v** in our notation. Thompson presents the information matrix in terms of direct products of the sub-matrices of **U** and **T**.

The model introduced implicitly assumes that $\mathfrak{d}$ is positive semi-definite, i.e. that any linear combination of the $\mathbf{u}_i$'s should have a non-negative variance. There is, however, no guarantee that the estimate $\hat{\mathfrak{d}}$ given by our procedure is positive semi-definite.

The situation is analogous to that of negative variance components in univariate data. There are several options open in univariate data (see Harville 1977) and these can be extended to deal with multivariate situations (for instance, Bock and Peterson 1975), but we have not investigated the merits of these different options.

4. Example

We illustrate the calculations of this method with the example data in Table 1. Usually male and female yearling weights have been analyzed separately ignoring the existence of a non-zero component of covariance. Theoretically, the sire correlation between male and female yearling weights should be close to one. In this example, t equals 2, and trait 1 refers to male yearling weights while trait 2 refers to female yearling weights. We begin the iterative procedure using the following initial estimates of $\mathfrak{d}$ and $\mathfrak{d}_e^2$.

$$\mathbf{B} = \begin{bmatrix} 254.00\,\mathbf{I} & 189.38\,\mathbf{I} \\ 189.38\,\mathbf{I} & 393.82\,\mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} .00614\,\mathbf{I} & -.00295\,\mathbf{I} \\ -.00295\,\mathbf{I} & .00396\,\mathbf{I} \end{bmatrix}$$

and

$$\tilde{\mathbf{R}}^{-1} = \begin{bmatrix} 2922.97\,\mathbf{I} & \mathbf{O} \\ \mathbf{O} & 1540.20\,\mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} 000342\,\mathbf{I} & \mathbf{O} \\ \mathbf{O} & .000649\,\mathbf{I} \end{bmatrix}$$

The complete mixed model equations (8) are of order 16, and are not shown, but (10) is shown below.

$$\mathbf{C}_{11} = \begin{bmatrix} 144.79 & 34.84 & 43.47 & 30.91 \\ 34.84 & 157.08 & 33.95 & 28.14 \\ 43.47 & 33.95 & 142.89 & 33.70 \\ 30.91 & 28.14 & 33.70 & 161.25 \end{bmatrix},$$

$$\mathbf{C}_{12} = \begin{bmatrix} 65.72 & 41.94 & 46.00 & 35.71 \\ 42.04 & 71.29 & 41.64 & 34.41 \\ 45.85 & 41.43 & 65.21 & 36.89 \\ 35.76 & 34.72 & 36.53 & 82.37 \end{bmatrix},$$

$$\mathbf{C}_{22} = \begin{bmatrix} 159.16 & 81.27 & 89.33 & 64.06 \\ 81.27 & 169.27 & 80.49 & 62.79 \\ 89.33 & 80.49 & 159.04 & 64.96 \\ 64.06 & 62.79 & 64.96 & 202.02 \end{bmatrix}.$$

The solutions to (9) are

$\hat{\mathbf{u}}_1' = (-3.836\ -13.830\ 14.379\ 3.288)$ and $\hat{\mathbf{u}}_2' = (11.548\ -23.942\ 13.344\ -.951)$. From these solution vectors and (11) we obtain the following:

$$\hat{\mathbf{v}}_1 = \alpha_{11}\hat{\mathbf{u}}_1 + \alpha_{12}\hat{\mathbf{u}}_2 \text{ and } \hat{\mathbf{v}}_2 = \alpha_{12}\hat{\mathbf{u}}_1 + \alpha_{22}\hat{\mathbf{u}}_2.$$

$$\hat{\mathbf{v}}_1' = (-.0576\ -.0142\ .0489\ .0230),$$

$$\hat{\mathbf{v}}_2' = (.0570\ -.0540\ .0104\ -.0135),$$

$$\text{and } \mathbf{f}' = (.003219\ -.001160\ .003226\ .006473\ .016190).$$

The matrices $\mathbf{U}$ and $\mathbf{T}$ in (14) are constructed from $\mathbf{C}$, $\mathbf{B}$, and $\mathbf{R}_*^{-1}$ and used to compute the expectations in (15)–(17). The resulting expectation matrix is

$$\mathbf{F} = \begin{bmatrix} .0622 & -.0453 & .0083 & .0063 & .0006 \\ & .1477 & -.0477 & -.0045 & -.0039 \\ & & .0695 & .0008 & .0059 \\ \text{symmetric} & & & .0217 & .0001 \\ & & & & .1039 \end{bmatrix} \times 10^{-4}.$$

Inverting $\mathbf{F}$ and solving for $\hat{\boldsymbol{\theta}}$ gives $\hat{\boldsymbol{\theta}}' = (288.07\ 188.71\ 393.26\ 2810.03\ 1539.49)$.

TABLE 2
Change in Variance and Covariance Component Estimates (kg^2) Through the Iteration Process.

Round	$\tilde{\sigma}_1^2$	$\tilde{\sigma}_{12}^2$	$\tilde{\sigma}_2^2$	$\tilde{\sigma}_{e1}^2$	$\tilde{\sigma}_{e2}^2$
0 (initial)	254.01	0.00	393.82	2922.97	1540.20
1	254.00	189.38	393.82	2922.97	1540.20
2	288.07	188.71	393.26	2910.03	1539.49
3	280.29	190.47	393.26	2913.83	1539.60
4	282.42	190.07	393.28	2912.86	1539.56
5	281.85	190.18	393.27	2913.12	1539.57
6	282.00	190.15	393.27	2913.05	1539.57

Table 2 shows the change in estimates through six rounds of iteration starting with an initial sire covariance of zero. This procedure converged rapidly for this example, but for larger data sets may require more rounds of iteration. Convergence will also depend on the adequacy of the initial estimates used to start the iterations.

From these results we could estimate the heritabilities of male and female yearling weights as $\hat{h}_1^2 = 4(282)/(282 + 2913) = .353$, $\hat{h}_2^2 = 4(393)/(393 + 1540) = .813$ and the sire correlation between these traits as $\hat{r}_{1,2} = 190/[(282)(393)]^{.5} = .571$.
This is just a numerical example and more data would be necessary before one could obtain reasonable estimates. Large numbers of observations would also be needed in order to use $\mathbf{F}^{-1}$ as the variance-covariance matrix of the estimates.

Résumé

On décrit une procédure d'estimation des composantes de la variance et de la covariance quand différentes variables sont mesurées sur différentes unités expérimentales. La procédure permet des modèles linéaires différents sur les différents traitements. La méthode consiste à maximiser la vraisemblance d'un ensemble de contrastes d'erreur. C'est une extension de la procédure présentée par THOMPSON (1973) dans le cas où toutes les variables sont mesurées sur tous les traitements. La méthode est de plus une version itérative de la procédure d'estimation quadratique non biaisée de norme minimum (MINQUE) de C.R. RAO (1971). Les calculs sont décrits à l'aide des solutions des Equations de HENDERSON (1973) sur le modèle mixte. La procédure a été développée pour estimer les composantes de la variance et de la covariance des pères à partir des poids de jeunes bovins males et femelles; un échantillon de ces données est utilisé pour illustrer la méthode.

References

Anderson, T. W. (1973). A symptotically efficient estimation of covariance matrices with linear structure. *Annals of Statistics 1*, 135–141.

Bock, R. D. and Peterson, A. C. (1975). A multivariate correction for attenuation. *Biometrika 62*, 673–678.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association 72*, 320–338.

Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics 9*, 226–252.

Henderson, C. R. (1973). Sire evaluation and genetic trends. *Proceedings of the Animal Breeding and Genetics Symposium in Honour of Dr. Jay L. Lush.* Published by the American Society of Animal Science and the American Dairy Science Association.

Kempthorne, O. (1969). *An Introduction to Genetic Statistics.* The Iowa State University Press, Ames, Iowa.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika 58,* 545–554.

Patterson, H. D. and Thompson, R. (1974). Maximum likelihood estimation of components of variance. *Proceedings of the 8th International Biometrics Conference,* Constanta, Romania, 197–207.

Rao, C. R. (1971). Minimum variance quadratic unbiased estimation of variance components. *Journal of Multivariate Analysis 1,* 445–456.

Searle, S. R. (1966). *Matrix Algebra for the Biological Sciences.* Wiley, New York.

Thompson, R. (1973). The estimation of variance and covariance components with an application when records are subject to culling. *Biometrics 29,* 527–550.

Received February 1976; *Revised May* 1977

Appendix

We derive here some of the terms used in Section 3.3. We let $\mathbf{S} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}^{-1}$ then the sums of squares of residuals $\gamma_i^2\hat{\mathbf{e}}_i'\hat{\mathbf{e}}_i$ can be written as

$$
\begin{aligned}
&(\mathbf{y} - \mathbf{X}\hat{\alpha} - \mathbf{Z}\hat{\mathbf{u}})'\mathbf{R}^{-1}\mathbf{D}_i\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\alpha} - \mathbf{Z}\hat{\mathbf{u}}) \\
&= (\mathbf{y} - \mathbf{Z}\hat{\mathbf{u}})'\mathbf{S}\mathbf{D}_i\mathbf{S}(\mathbf{y} - \mathbf{Z}\mathbf{u}) \text{ (using (5.2) of Harville 1977),} \\
&= \mathbf{y}'\mathbf{S}\mathbf{D}_i\mathbf{S}\mathbf{y} - 2\mathbf{y}'\mathbf{S}\mathbf{D}_i\mathbf{S}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{S}\mathbf{y} + \mathbf{y}'\mathbf{S}\mathbf{A}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1} \\
&\cdot (\mathbf{Z}'\mathbf{S}\mathbf{D}_i\mathbf{S}\mathbf{Z} + (\mathbf{G}^{-1} - \mathbf{G}^{-1})\mathbf{D}_{ii})(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{S}\mathbf{y} \text{ (using (6)),} \\
&= \gamma_i^2\mathbf{y}_i'\mathbf{P}_i\mathbf{y}_i - \gamma_i^2\hat{\mathbf{u}}_i'\mathbf{Z}_i'\mathbf{P}_i\mathbf{y}_i - \sum_{j=1}^{t}(\alpha_{ij}\gamma_i)\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_j \text{ (using } \mathbf{S}\mathbf{S} = \mathbf{R}^{-1}\mathbf{S}). \qquad \text{(A1)}
\end{aligned}
$$

The terms of the information matrix (or equivalently the coefficient of θ_n in the expectation of f_m) are $1/2\text{tr}(\mathbf{M}\mathbf{V}_m\mathbf{M}\mathbf{V}_n)$ where $\mathbf{M} = \mathbf{S} - \mathbf{S}\mathbf{Z}(\mathbf{G}^{-1} + \mathbf{Z}'\mathbf{S}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{S}$ (section 5 and (3.7) of Harville). If θ_m and θ_n represent σ_{ij} and σ_{kl}, then

$$tr(\mathbf{M}\mathbf{V}_m\mathbf{M}\mathbf{V}_n) = \text{tr}[\mathbf{U}(\mathbf{I} * \mathbf{D}_{ij})\mathbf{U}(\mathbf{I} * \mathbf{D}_{kl})], \qquad \text{(A2)}$$

(using $\mathbf{Z}'\mathbf{M}\mathbf{Z} = \mathbf{G}^{-1} - \mathbf{G}^{-1}(\mathbf{G}^{-1} + \mathbf{Z}'\mathbf{S}\mathbf{Z})^{-1}\mathbf{G}^{-1} = \mathbf{U}$). Using straightforward but tedious algebra, similar to that used in the derivation of (A1), we find that if θ_m and θ_n represent σ_{ei}^2 and σ_{kl} then

$$tr(\mathbf{M}\mathbf{V}_m\mathbf{M}\mathbf{V}_n) = tr[\mathbf{U}(\mathbf{I} * \mathbf{D}_{ii})\mathbf{T}(\mathbf{I} * \mathbf{D}_{kl})], \qquad \text{(A3)}$$

and if θ_m and θ_n represent σ_{ei}^2 and σ_{ej}^2 then

$$tr(\mathbf{M}\mathbf{V}_m\mathbf{M}\mathbf{V}_n) = (N_i - p_i - s)\gamma_i\delta_{ij} + tr[\mathbf{T}(\mathbf{I} * \mathbf{D}_{ii})\mathbf{T}(\mathbf{I} * \mathbf{D}_{jj})]. \qquad \text{(A4)}$$

7

A note on the estimation of the parameters of the autoregressive-moving average process

by

D.M. Cooper and R. Thompson

Biometrika (1977), **64**, 3, *pp.* 625–8

A note on the estimation of the parameters of the autoregressive-moving average process

BY D. M. COOPER* AND R. THOMPSON

Agricultural Research Council Unit of Statistics, Edinburgh

SUMMARY

Patterson & Thompson (1971) have described a modified maximum likelihood technique for estimating variance components in the additive mixed model. We consider the use of this method in estimating the parameters of the autoregressive-moving average model, and examine its performance relative to some established techniques for 200 simulations of a first-order moving average process.

Some key words: Autoregressive-moving average model; Modified maximum likelihood; Simulation.

1. INTRODUCTION

Suppose a time series W_t is generated by the autoregressive moving average process

$$W_t - \phi_1 W_{t-1} - \ldots - \phi_p W_{t-p} = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}, \tag{1·1}$$

where a_t is a sequence of independent normal random variables having mean zero and variance σ^2. The joint density for $W = (W_1, \ldots, W_n)'$ may be written

$$f(W|\phi, \theta, \sigma^2) = |V|^{-\frac{1}{2}} (2\pi\sigma^2)^{-\frac{1}{2}n} \exp(-\tfrac{1}{2} W' V^{-1} W/\sigma^2), \tag{1·2}$$

where $\phi = (\phi_1, \ldots, \phi_p)'$, $\theta = (\theta_1, \ldots, \theta_q)'$ and $V = ((v_{ij}))$ is the variance–covariance matrix of W. The v_{ij} are functions of ϕ and θ and are such that $v_{ij} = v_{kl}$ if $|i-j| = |k-l|$. If now, instead of taking $E(W) = 0$, we let $E(W) = XA$, where X is a known $n \times m$ matrix of rank $m < n$ and A is an $m \times 1$ matrix of parameters, we have the joint density

$$f(W|\phi, \theta, A, \sigma^2) = |V|^{-\frac{1}{2}} (2\pi\sigma^2)^{-\frac{1}{2}n} \exp\{-\tfrac{1}{2}(W - XA)' V^{-1} (W - XA)/\sigma^2\}. \tag{1·3}$$

We are interested in estimating the parameters ϕ and θ of (1·3). Least squares estimates are found by minimizing $(W - XA)' V^{-1} (W - XA)$ with respect to ϕ and θ, after replacing A by its least squares estimator $(X' V^{-1} X)^{-1} X' V^{-1} W$. Maximum likelihood estimates are found by maximizing (1·3) after replacing A and σ^2 by their maximum likelihood estimators

$$(X' V^{-1} X)^{-1} X' V^{-1} W, \quad W'\{V^{-1} - V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1}\} W/n.$$

For both methods of estimation a function minimization routine is used. Expressions for V^{-1} and $|V|$ have been given by Newbold (1974). In his notation,

$$V^{-1} = L_n' \{I - X_n (X_n' X_n + \Omega^{-1})^{-1} X_n'\} L_n, \quad |V| = |\Omega|\, |X_n' X_n + \Omega^{-1}|.$$

Here L_n, X_n and Ω are respectively $n \times n$, $n \times (p+q)$ and $(p+q) \times (p+q)$ matrices whose elements are functions of ϕ and θ. These matrices are readily calculated for particular values of p and q. Alternative derivations of V^{-1} and $|V|$ give slightly different expressions for these quantities. In our calculations we have used one such pair of alternatives.

Having outlined the two most commonly advocated techniques of estimation of ϕ and θ, we go on to describe the method of modified maximum likelihood estimation of the parameters of (1·3).

2. MODIFIED MAXIMUM LIKELIHOOD ESTIMATION

Patterson & Thompson, when estimating variance components, consider a model of the form (1·3). Their V has a different structure from that being considered here, but has elements which are functions of parameters which are to be estimated. They consider expressing (1·3),

* Present address: Department of Statistics, University College of Wales, Aberystwyth.

regarded as a likelihood, as the product of two likelihoods associated with transformations SW and QW, where S and Q satisfy certain orthogonality properties. Here Q is chosen so that QX is of rank m with linear functions of QW exactly determining linear functions of the elements of A, leaving no degrees of freedom for further parameter estimation. The likelihood associated with QW thus contains no information on ϕ and θ, which are estimated from the remaining section of the likelihood. Choosing

$$S = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X^1V^{-1}, \quad Q = X'V^{-1},$$

that part of the likelihood (1·3) associated with SW is

$$(2\pi\sigma^2)^{-\frac{1}{2}(n-m)}(l_1 \dots l_{n-m})^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}W'S(SVS)^{-}SW/\sigma^2\}\left|\frac{dSW}{dW}\right|. \tag{2·1}$$

Here SVS is singular of rank $n-m$ and $l_1, \dots, l_{n-m}$ are its nonzero eigenvalues. Now $SVS = S$, so V is a generalized inverse of S and hence of SVS, so that (2·1) can be written

$$(2\pi\sigma^2)^{-\frac{1}{2}(n-m)}(l_1 \dots l_{n-m})^{\frac{1}{2}} \exp(-\tfrac{1}{2}W'SW/\sigma^2),$$

because the Jacobian in (2·1) is $(l_1 \dots l_{n-m})$. Let P be the $n \times (n-m)$ matrix of orthonormalized eigenvectors of S which are associated with its nonzero eigenvalues. Then $P'X = 0$ and

$$|P'\{V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}\}P| = \prod_{i=1}^{n-m} l_i,$$

where l_i are as previously defined. Now

$$|P'\{V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}\}P|\,|X'V^{-1}X| = \begin{vmatrix} P'V^{-1}P & P'V^{-1}X \\ X'V^{-1}P & X'V^{-1}X \end{vmatrix}. \tag{2·2}$$

By manipulating the right-hand side of (2·2), and using the properties $P'X = 0$ and $|P'P| = 1$, we find

$$\prod_{i=1}^{n-m} l_i = |V^{-1}|\,|X'X|\,|X'V^{-1}X|^{-1}.$$

The modified likelihood associated with SW is therefore

$$(2\pi\sigma^2)^{-\frac{1}{2}(n-m)}|V|^{-\frac{1}{2}}|X'X|^{\frac{1}{2}}|X'V^{-1}X|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}W'\{V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}\}W/\sigma^2]. \tag{2·3}$$

This expression has been given by Harville (1974) whose derivation includes integration over A. Modified maximum likelihood estimates are found by maximizing (2·3) numerically with respect to ϕ and θ, after replacing σ^2 by its modified maximum likelihood estimator $W'SW/(n-m)$. Then A is estimated by $(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}W$, where $\hat{V}$ is the variance–covariance matrix of W with ϕ and θ replaced by their modified maximum likelihood estimates.

3. Results of a simulation study of the first-order moving average process

The method of the foregoing section was considered in an attempt to produce improved methods of parameter estimation for a set of 200 simulations of a moving average process for which least squares estimation appeared unsatisfactory. The simulated series were generated by the process

$$W_t - 0{\cdot}018751 = a_t - 0{\cdot}7539a_{t-1} \quad (t = 1, \dots, 59), \tag{3·1}$$

where a_t is a sequence of independent normal random variables having mean zero and variance 0·0549. In the notation of the previous section X is an n-vector of ones and $A = \mu = 0{\cdot}018751$. Two notable features of the first-order moving average model (3·1) are the presence of a mean, μ, and the closeness of the moving average parameter $\theta = 0{\cdot}7539$ to the unit circle. Frequency distributions of estimates of the moving average parameter θ are given in Table 1.

Table 1. *First-order moving average process. Frequency distributions of estimates* $\hat{\theta}$, *using* (*a*) *least squares with correct backforecast* a_0, (*b*) *least squares with incorrect backforecast* a_0, (*c*) *maximum likelihood with estimation of the mean*, (*d*) *modified maximum likelihood*, *and* (*e*) *maximum likelihood with the mean set to its known value*

θ	*a*	*b*	*c*	*d*	*e*	θ	*a*	*b*	*c*	*d*	*e*
$(-\infty, 0{\cdot}45)$	1	1	1	1	2	[0·80, 0·85)	28	28	32	30	33
[0·45, 0·50)	1	1	2	4	4	[0·85, 0·90)	16	19	17	22	20
[0·50, 0·55)	2	2	1	1	1	[0·90, 0·95)	2	11	2	8	12
[0·55, 0·60)	2	2	3	11	7	[0·95, 1·00)	0	57	0	1	2
[0·60, 0·65)	5	5	11	12	14	1·00	82	—	65	19	8
[0·65, 0·70)	12	12	11	12	17	(1·00, 1·05)	—	12	—	—	—
[0·70, 0·75)	18	18	23	43	36	[1·05, 1·10)	—	1	—	—	—
[0·75, 0·80)	31	31	32	36	44	Total	200	200	200	200	200

First we consider least squares estimation. The form of the sum of squares surface $(W-XA)'\,V^{-1}(W-XA)$ for the first-order moving average model with $E(W)=0$, has been noted by Box & Jenkins (1970, §A.7.6). There are global minima of zero as $\theta \to \pm\infty$, and usually a local minimum in the θ interval $[-1, 1]$. Least squares estimation aims to find this local minimum if it exists. Box & Jenkins suggest using a function minimization routine to find estimates of μ and θ, calculating the a_t's of (3·1) recursively, for different parameter values, and using a 'backforecast' estimator for a_0. The use of the correct estimator leads to estimates of μ and θ at which the sum of squares function is minimized. By choosing suitable initial estimates of these parameters in the function minimization routine used, the local minimum, in the θ interval $[-1, 1]$ will be found, if it exists. Box & Jenkins also suggest using an approximation to the backforecast estimator of a_0. The results of our simulation study have shown that this approximation can give misleading estimates of θ. Its use distorts the sum of squares function for θ values near 1 and -1 in such a way as to create a minimum with respect to θ near these values when none exists in the correct sum of squares function. This leads to spurious estimates of θ. We believe this point is worth noting because the time series estimation program USES written by ISCOL has produced these misleading estimates when applied to some of our simulated series.

In estimating the parameters of the 200 simulations used, it was found that the sum of squares surfaces with μ replaced by the generalized least squares estimator,

$$(X'V^{-1}X)^{-1}X'V^{-1}W,$$

had no local minimum in the θ interval $[-1, 1]$ for 82 simulations. When the approximate backforecast estimator of a_0 was used, in each of the 82 cases, a minimum of the approximate sum of squares surface at a point between $\theta = 0{\cdot}85$ and $\theta = 1{\cdot}1$ was found. Since estimates of θ are constrained, by consideration of the nature of the model used, to lie between -1 and 1, in the cases where no local minimum was found, 1 was taken to be the least squares estimate of θ.

Secondly, maximum likelihood estimation of μ and θ was examined. The likelihood surface represented by (1·3) with μ and σ^2 replaced by their maximum likelihood estimates, is symmetric in θ and $1/\theta$, and always has a maximum in the interval $[-1, 1]$ with nonzero probability of estimates at one or other of the values -1 and 1. Of the maximum likelihood estimates 65 out of 200 were equal to 1. When μ was set to its true value in (1·3) 8 estimates were equal to 1.

Lastly, estimates of θ were found using the modified maximum likelihood technique. This gave 19 estimates equal to 1.

We note that, for the first-order moving average model, when $\theta = 0{\cdot}75$ the theoretical autocorrelation at lag 1 is 0·48. The maximum value possible for any real model with zero autocorrelation for higher lags is 0·50. It is of interest to relate these estimates to the recently introduced minimum norm quadratic unbiased estimates, MINQUE, of Rao (1973, pp. 303–5).

With a norm based on $\theta = 1$ and a linear reparameterization of the elements of the variance–covariance matrix it was found for the 200 simulations that some of the MINQUE estimates of the lag 1 autocorrelation were greater than 0·5 and these corresponded to estimates of θ which were equal to 1 using modified maximum likelihood estimation. This is not altogether surprising because of the connexion between modified maximum likelihood estimation and an iterated form of MINQUE (Patterson & Thompson, 1975; Hocking & Kutner 1975).

We are grateful to Mr P. R. Fisk of the University of Edinburgh for providing 200 simulated series for analysis and for his advice and suggestions concerning this work.

References

Box, G. E. P. & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden Day.

Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* **61**, 383–5.

Hocking, R. R. & Kutner, M. H. (1975). Some analytical and numerical comparisons of estimators for the mixed analysis of variance model. *Biometrics* **31**, 19–28.

Newbold, P. (1974). The exact likelihood function for a mixed autoregressive moving average process. *Biometrika* **61**, 423–6.

Patterson, H. D. & Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika* **58**, 545–54.

Patterson, H. D. & Thompson, R. (1975). Maximum likelihood estimation of components of variance. In *Proc. 8th Int. Biometric Conf.*, Ed. L. C. A. Corsten and T. Postelnicu, pp. 197–208. Bucharest: Academy of the Socialist Republic of Romania.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd edition. New York: Wiley.

[*Received October* 1976. *Revised March* 1977]

8

A note on restricted maximum likelihood estimation with an alternative outlier model

by

R. Thompson

A Note on Restricted Maximum Likelihood Estimation with an Alternative Outlier Model

BY

R. THOMPSON

Reprinted from

JOURNAL OF THE ROYAL STATISTICAL SOCIETY

SERIES B (METHODOLOGICAL)

Volume 47, No. 1, 1985

(pp. 53–55)

PRINTED FOR PRIVATE CIRCULATION

1985

J. R. Statist. Soc. B, (1985)
47, *No.* 1, *pp.* 53–55

A Note on Restricted Maximum Likelihood Estimation with an Alternative Outlier Model

By R. THOMPSON

ARC Unit of Statistics, Edinburgh, Scotland

[Received May 1983. Revised November 1983]

SUMMARY

For single outliers in normal theory fixed effects models a mean slippage model is commonly used. An alternative is to model the outlier as arising from an unknown observation with inflated variance. Maximum likelihood estimates for the position of the outlier under the two models need not agree. This paper considers maximizing a restricted part of the likelihood to estimate the variance parameters and characterizes these estimates in terms of standard least squares parameters. It is shown that the residual variance and outlier position are the same under both models.

Keywords: OUTLIERS; LINEAR MODELS; RESTRICTED MAXIMUM LIKELIHOOD ESTIMATION; MEAN SLIPPAGE MODELS; VARIANCE SLIPPAGE MODELS

1. INTRODUCTION

In a recent paper Cook, Holschuh and Weisberg (1982) consider two models for single outliers in fixed effects linear models. One is based on the assumption that contamination gives rise to slippage in the expected values of the observations (Weisberg, 1980, Section 5.3). Cook *et al.* point out the key role of Studentized residuals in this model, for instance in estimating the position of an outlier or testing for the presence of an outlier.

Alternatively one can assume that an outlier arises from an error term with an increased variance. Cook *et al.* argue intuitively that it seems reasonable that Studentized residuals might play a similar role under this alternative model but show that maximum likelihood estimates of outlier position can differ under these two models.

Cook *et al.* note that their models can be fitted into a linear model framework discussed by Harville (1977). They do not note that Harville recommends a restricted maximum likelihood (*REML*) approach using only a restricted part of the likelihood to estimate the variance parameters. Patterson and Thompson (1971) noted that this *REML* approach takes account of loss of degrees of freedom in estimating fixed effects. In this note, *REML* estimates for the alternative model are derived and expressed in standard least squares statistics. It is shown that the observation having the largest Studentized residual is the one picked out as the outlier.

2. THE MODEL

Using, as far as is possible (vectors and matrices are not now printed in bold type), the notation of Cook *et al.* then the observation vector y of length n is assumed to have expectation $X\beta$ where X is a $n \times p$ full rank matrix, and β is a vector of p unknowns. The variance of all observations, except the ith, is assumed to be σ^2, and the ith has variance $w\sigma^2 (w \geqslant 1)$. The variance matrix of y, when the ith observation has an inflated variance $w\sigma^2$ will be denoted by W_i. It is required to estimate β, σ^2, w and i.

It is useful to define $X' = (x_1, x_2, \ldots, x_n)'$ with x_i a vector of length p and also $v_i = x_i'(X'X)^{-1}x_i$. Also let $\tilde{y}_j = x_j'\tilde{\beta}$, a predicted value for the jth observation using $\tilde{\beta} = (X'X)^{-1}X'y$, let $(n-p)\,\hat{\sigma}^2 = \Sigma_j(\tilde{y}_j - y_j)^2$ and, provided $v_i < 1$, define the ith Studentized

Present address: Animal Breeding Research Organisation, King's Buildings, Mayfield Road, Edinburgh EH9 3 JZ, Scotland.

© 1985 Royal Statistical Society 0035–9246/85/47053 $2.00

residual as $t_i = (y_i - \tilde{y}_i)/(\hat{\sigma}(1-v_i)^{\frac{1}{2}})$. This estimator of the variance, $\hat{\sigma}^2$, differs from that of Cook *et al.*, $\tilde{\sigma}^2$, being larger by a factor $\{n/(n-p)\}$. However the definition of Studentized residuals is equivalent to that of Cook *et al.*

3. RESTRICTED MAXIMUM LIKELIHOOD ESTIMATION

To find estimates of the parameters, the value of i is fixed and estimates of σ^2 and w, σ_i^2 and w_i, are found by maximizing a restricted likelihood $l_i'(\sigma^2, w; y)$. The procedure is repeated for the n values of i and the chosen value of i is that giving the largest value of $l_i'(\sigma_i^2, w_i; y)$. The log-likelihood used is not that of all the data, but only error contrasts i.e. contrasts $c'y$ such that $E(c'y) = 0$ (Patterson and Thompson, 1971). Then if the ith observation is inflated the restricted log-likelihood can be found from (Harville, 1977)

$$2l_i'(\sigma^2, w; y) = -n \log \sigma^2 - \log w - \log [\det(X' W_i^{-1} X)] - (y - X\beta_i(w))' W_i^{-1} (y - X\beta_i(w))/\sigma^2,$$

where

$$\begin{aligned} \beta_i(w) &= (X' W_i^{-1} X)^{-1} (X' W_i^{-1} y) \\ &= \tilde{\beta} - (y_i - \tilde{y}_i) [(w-1)/(w-(w-1)v_i)] (X'X)^{-1} x_i, \end{aligned} \tag{1}$$

which relates the weighted estimate to the unweighted estimate (Cook *et al.*, 1982). Hence

$$\begin{aligned} &(y - X\beta_i(w))' W_i^{-1} (y - X\beta_i(w)) \\ &= \Sigma (y_j - \tilde{y}_j)^2 - (w-1)(y_i - \tilde{y}_i)^2/(w-(w-1)v_i) \\ &= \Sigma (y_j - \tilde{y}_j)^2 - (y_i - \tilde{y}_i)^2/(1-v_i) + (y_i - \tilde{y}_i)^2/\{(1-v_i)(w-(w-1)v_i)\} \\ &= R_{[i]} + (y_i - \tilde{y}_i)^2/(1-v_i)(w-(w-1)v_i), \end{aligned}$$

where $R_{[i]}$ is the residual sum of squares when the ith observation is ignored (or equivalently as $w \to \infty$).

In a similar way, using $\det(A + x_i' x_i) = \det(A)(1 + x' A^{-1} x_i)$ it can be shown that, apart from an additive constant,

$$\begin{aligned} 2l_i'(\sigma^2, w; y) = &-(n-p-1) \log \sigma^2 - \log (w-(w-1)v_i)\sigma^2 \\ &-[R_{[i]} + (y_i - \tilde{y}_i)^2/\{(1-v_i)(w-(w-1)v_i)\}]/\sigma^2. \end{aligned}$$

The sums of squares in this expression, $R_{[i]}$ and $(y_i - \tilde{y}_i)^2/(1-v_i)$, are equivalent to the terms used by Cook *et al.* in their likelihood and $(n-p-1)\sigma^2$ and $(w-(w-1)v_i)\sigma^2$ are the expectations of the two sums of squares. By differentiating this likelihood with respect to σ^2 and $(w-(w-1)v_i)\sigma^2$ the maximum value of $l_i'(\sigma^2, w; y)$ is seen to occur when

$$\sigma^2 = R_{[i]}/(n-p-1) = (n-p-t_i^2)\hat{\sigma}^2/(n-p-1) \tag{2}$$

and

$$(w-(w-1)v_i)\sigma^2 = t_i^2 \hat{\sigma}^2$$

or

$$w = [(n-p)(t_i^2 - v_i) - t_i^2 (1-v_i)]/[(n-p)(1-v_i) - t_i^2 (1-v_i)]. \tag{3}$$

The model only allows inflation of variance and w in (3) is < 1 if and only if $t_i^2 < 1$. Hence estimates σ_i^2 and w_i satisfy (2) and (3) if $t_i^2 \geqslant 1$ and satisfy $\sigma_i^2 = \hat{\sigma}^2$, $w_i = 1$ if $t_i^2 < 1$. $2l_i'(\sigma_i^2, w_i, y)$ is then $-(n-p-1)\log(n-p-t_i^2) - \log(t_i^2) - (n-p)\log\hat{\sigma}^2 + (n-p-1)\log(n-p-1) - (n-p)$ if $t_i^2 \geqslant 1$ and $-(n-p)\log\hat{\sigma}^2 - (n-p)$ if $t_i^2 < 1$. Note that (2) also gives the *REML* estimate of variance if a mean slippage model for the ith observation is assumed.

As $(n-p-1)\log(n-p-t_i^2) + \log(t_i^2)$ decreases as t_i^2 increases from 1 to $n-p$, and

$t_i^2 = (n-p)$ is an upper bound on t_i^2 (if $t_i^2 > (n-p)$ then $R_{[i]}$ would be less than zero) then the maximum value of $l_i'(\sigma_i^2, w_i; y)$ will correspond to the largest absolute value of the Studentized statistic. This suggests as an *REML* estimation procedure (i) finding the largest t_i^2 value, say $i = k$ (ii) finding estimates of $\sigma_k^2\, w_k$ from (2) and (3), (iii) finding β_k (w_k) from (1).

As an example, the data given by Cook *et al.* were used to fit a model $y_i = \alpha + \gamma x_i$ with the ith observation having an inflated variance w_i. For eight observations, those with Studentized residuals less than one, the maximum value of $l_i'(\sigma_i^2, w_i; y)$ was -29.388 with the estimate of w_i being one. The observations 7, 9 and 11, with Studentized residuals 1.41, 7.39 and 8.48 give rise to estimates of inflated variances of 1.41, 7.39 and 8.48 with maximum values of the restricted likelihood being $-29{,}384$, -28.120 and -28.086. The overall maximum restricted likelihood therefore occurs when the inflated variance is associated with the observation with the largest absolute Studentized residual. The *REML* estimators are then $\hat{\alpha} = 4.66$, $\hat{\gamma} = 0.87$, $\hat{\sigma}^2 = 152.45$, $\hat{w} = 8.48$, $i = 11$.

REFERENCES

Cook, R. D., Holschuh, N. and Weisberg, S. (1982) A note on an alternative outlier model. *J. R. Statist. Soc.* B, **44**, 370–376.

Harville, D. A. (1977) Maximum likelihood approaches to variance component estimation and to related problems (with Discussion). *J. Amer. Statist. Ass.*, **72**, 320–340.

Patterson, H. D. and Thompson, R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545–554.

Weisberg, S. (1980) *Applied Linear Regression*. New York: Wiley.

ROYAL STATISTICAL SOCIETY
25 Enford Street, London, W1H 2BH

9

A note on the W transformation

by

R. Thompson

A Note on the W Transformation

R. Thompson

Agricultural Research Council
Unit of Statistics
21 Buccleuch Place
Edinburgh EH8 9LN, Scotland

A matrix transformation, named the W transformation by Hemmerle and Hartley [3], is useful in the maximum likelihood estimation of a mixed analysis of variance model. A simpler computational form for W is suggested.

KEY WORDS

Mixed A.O.V. Model
Variance Components
Maximum Likelihood
Computational Methods

Hemmerle and Hartley [3] discuss the computational problems of maximum likelihood for the general mixed analysis of variance model. They suggest that the computation can be reduced (compared with other methods studied by Hartley and his co-workers [1], [2]) by use of a matrix transformation called by them the W transformation.

The model considered in [3] is of the form

$$y = X\alpha + \epsilon \tag{1}$$

where ϵ has a multivariate normal distribution with variance-covariance matrix

$$\sigma^2 H = \sigma^2\left(I_n + \sum_{i=1}^{c} \gamma_i U_i U_i'\right) \tag{2}$$

X and U_i are known $n \times k$ and $n \times m_i$ matrices respectively, y and α are vectors with n and k elements respectively. Let

$$m = \sum_{i=1}^{c} m_i$$

and

$$V = [U_1 \mid U_2 \cdots \mid U_c].$$

We write W in two forms, one as defined in [3] in terms of submatrices $W(i, j)$,

$$W = \begin{bmatrix} W(1,1) & \cdots & W(1,c) & W(1,c+1) & W(1,c+2) \\ \vdots & & \vdots & \vdots & \vdots \\ W(c,1) & \cdots & W(c,c) & W(c,c+1) & W(c,c+2) \\ W(c+1,1) & \cdots & W(c+1,c) & W(c+1,c+1) & W(c+1,c+2) \\ W(c+2,1) & \cdots & W(c+2,c) & W(c+2,c+1) & W(c+2,c+2) \end{bmatrix}. \tag{3}$$

where $W(i, j)$ is a $(m_i \times m_j)$ matrix. For convenience we let $m_{c+1} = k$ and $m_{c+2} = 1$. Another conveinent partition is

$$W = \begin{bmatrix} W_{11} & W_{12} & W_{13} \\ W_{21} & W_{22} & W_{23} \\ W_{31} & W_{32} & W_{33} \end{bmatrix} \tag{4}$$

We write the submatrices of (4) in three forms, the first as defined in [3] the second as it was suggested they be computed and thirdly, for three of them, an alternative form suggested by Patterson and Thompson in a similar situation [4]. It is probably worth commenting that although a matrix ΔW was introduced to help in the derivation of the second form in [3], it was not used in the iterative procedure outlined in that paper. The partition in (4) corresponds to classifying the rows and columns of W according to whether they are associated with fixed effects, random effects or the y variate.

The $W(i, j)$ in [3] are submatrices of W_{ij} and are therefore defined by (5)–(10).

		Dimensions of Sub-Matrix	
$W_{11} = V'H^{-1}V = V'V - V'VQ^{-1}V'V = D^{-1} - D^{-1}Q^{-1}D^{-1}$		$(m \times m)$	(5)
$W_{21} = X'H^{-1}V = X'V - X'V\,Q^{-1}V'V =$	$X'VQ^{-1}D^{-1}$	$(k \times m)$	(6)
$W_{31} = y'H^{-1}V = y'V - y'VQ^{-1}V'V =$	$y'VQ^{-1}D^{-1}$	$(1 \times m)$	(7)
$W_{22} = X'H^{-1}X = X'X - X'VQ^{-1}V'X$		$(k \times k)$	(8)
$W_{32} = y'H^{-1}X = y'X - y'VQ^{-1}V'X$		$(1 \times k)$	(9)
$W_{33} = y'H^{-1}y = y'y - y'VQ^{-1}V'y$		(1×1)	(10)

Since W is symmetric $W_{12} = W_{21}'$ etc.

Received February 1975; revised July 1975

The matrix D is the $m \times m$ diagonal matrix given by

$$D = \begin{bmatrix} \gamma_1 I_{m_1} & & & \\ & \gamma_2 I_{m_2} & & \\ & & \ddots & \\ & & & \gamma_c I_{m_c} \end{bmatrix}$$

where I_{m_i} is the identity matrix of order m_i . Q is given by

$$Q = (D^{-1} + V'V).$$

The advantage of the alternative form is that, for example in (5), we replace the multiplication of 3 square symmetric matrices which requires $m^2(3m + 1)/2$ multiplications by the multiplication of a symmetric matrix by two diagonal matrices which requires $m(m + 1)$ multiplications.

In fact, if we define $W^*(i, j) = \gamma_i\gamma_j W(i, j)$ $(i, j = 1, \cdots, c + 2)$ (where $\gamma_{c+1} = \gamma_{c+2} = 1$ for convenience) then we can avoid altogether the multiplication by the diagonal matrix D. Then

$$W_{11}^* = D - Q^{-1} \tag{11}$$

$$W_{21}^* = X'VQ^{-1} \tag{12}$$

$$W_{31}^* = y'VQ^{-1} \tag{13}$$

and

$$W_{22}^* = W_{22}\,, \quad W_{32}^* = W_{32} \quad \text{and} \quad W_{33}^* = W_{33}\,. \tag{14}$$

The equations used in [3] for estimation can be written down just as easily in terms of W^* as W. For instance equation (51) in [3] is

$$\frac{\delta\lambda}{\delta\gamma_i} = -\tfrac{1}{2}\,\mathrm{Tr}\,[W(i, i)] + \frac{1}{2\hat{\sigma}^2}\,P(i)'P(i) \tag{15}$$

with $P(i)' = W(c + 2, i) - \alpha W(c + 1, i)$ or equivalently

$$\frac{\delta\lambda}{\delta\gamma_i} = \gamma_i^{-2}\left[-\tfrac{1}{2}\,\mathrm{Tr}\,[W^*(i, i)] + \frac{1}{2\hat{\sigma}^2}\,P^*(i)'P^*(i)\right] \tag{16}$$

with $P^*(i)' = W^*(c + 2, i) - \alpha W^*(c + 1, i)$.

The number of multiplications saved per iteration by using W^* rather than W is of the order of $m^2(3m + 2k + 3)/2 - c(c + 4)/2$. The first term represents the saving in evaluating (11)–(13) instead of (5)–(7) (using the formulae in [3]) and the second represents the loss from evaluating terms such as (16) rather than as (15). Since c, the number of variance components, is usually small compared with m, the number of random effects, the second term is usually negligible compared with the first.

As Hemmerle and Hartley note, most of the computation in each iteration is in the formation of W, especially if m is large. By using W^*, instead of W, we can save three quarters of the computational effort of forming W, when the number of fixed effects is small compared to the number of random effects, m, and five ninths when $m = k$. The savings by using the W^* transformation therefore seem worthwhile.

References

[1] Hartley, H. O., and Rao, J. N. K. (1967). "Maximum Likelihood Estimation for the Mixed Analysis of Variance Model", *Biometrika*, 54, 93–108.

[2] Hartley, H. O. and Vaughn, W. K. (1972). "A Computer Program for the Mixed Analysis of Variance Model Based on Maximum Likelihood," In *Statistical Papers in Honour of George W. Snedecor*, Ed. by T. A. Bancroft, 129–144. Ames, Iowa State University Press.

[3] Hemmerle, W. J. and Hartley, H. O. (1973). "Computing Maximum Likelihood for the Mixed A.O.V. Model Using the W Transformation," *Technometrics*, 15, 819–31.

[4] Patterson, H. D. and Thompson, R. (1971). "Recovery of Inter-block Information when Block Sizes are Unequal," *Biometrika*, *58*, 545–54.

10

Extimation of variance components : what is missing in the EM algorithm?

by

R. Thompson and K. Meyer

J. Statist. Comput. Simul., 1986, Vol. 24, pp. 215–230
0094-9655/86/2404-0215 $18.50/0
© 1986 Gordon and Breach, Science Publishers, Inc.
Printed in Great Britain

Estimation of Variance Components: What is Missing in the *EM* Algorithm?

R. THOMPSON and K. MEYER

Animal Breeding Research Organisation and Institute of Animal Genetics, Kings Buildings, Edinburgh, Scotland

(*Received June 4, 1985*)

The *EM* algorithm is a frequently advocated algorithm for the estimation of variance components. A faster converging algorithm is developed using alternative parameterizations based on the analysis of variance. The procedure is exemplified using designs with two and three variance components and with multivariate designs using parameter values relevant to animal breeding data.

KEY WORDS: Variance components, *EM* algorithm, *REML*, analysis of variance.

1. THE MODEL

We consider estimation in linear models specified by

$$\mathbf{y}=\mathbf{X}\boldsymbol{\alpha}+\sum_{i=1}^{c}\mathbf{Z}_i\mathbf{b}_i+\mathbf{e} \tag{1.1}$$

where $\mathbf{y}$ is a $n\times 1$ vector of observed responses, $\mathbf{X}$ and $\mathbf{Z}_i$ $(i=1,\ldots,c)$ are known matrices of size $n\times p$ and $n\times q_i$, α is a $p\times 1$ vector of fixed effects, and each $\mathbf{b}_i$ is a $q_i\times 1$ vector of random effects distributed independently as $N(0,\sigma_i^2\mathbf{I})$ and $\mathbf{e}$ is a $n\times 1$ vector of error

terms distributed as $N(0, \sigma^2\mathbf{I})$. Suppose elements of $\mathbf{Z}_i$ have values zero or one and if each row of $\mathbf{Z}_i$ has only one non-zero element. Eq. (1.1) can then be thought of as an ANOVA mixed model.

We consider in this paper computation of variance components for this model by restricted maximum likelihood (*REML*) a procedure based on maximizing the likelihood (L) of error contrasts (Patterson and Thompson, 1971). If the data structures are unbalanced, iterative schemes are needed. We have found that iterative schemes using the expected values of the second differentials of L (Harville (1977), Patterson and Thompson (1971)) are often computationally feasible and can converge fairly quickly. In some cases, results on partitioned matrices (Thompson 1977a), Robinson, Thompson and Digby (1982)) can give a more efficient computing strategy. But there are data structures, for example Meyer (1983), where the calculation of second differentials is computationally expensive.

An alternative approach, for example Dempster, Selwyn, Patel and Roth (1984) and Laird (1982) is to use the *EM* algorithm. This is a general purpose algorithm, discussed by Dempster, Rubin and Laird (1977), that can be used to compute maximum likelihood estimates for a large number of statistical models, that can be thought of as using the first differentials of L. Thompson (1977b) used asymptotic formulae of Dempster *et al.* (1977) to show theoretically that the algorithm can be slow to converge for a one-way classification, especially for values of variance components common in animal breeding studies. This slow convergence has been found in other designs, for instance Thompson (1979).

However for most balanced designs there is a unique partition of the sum of squares into orthogonal parts due to the various factors in the model. It is then a simple matter to transform the mean squares for random factors, which estimate stratum variances (Nelder, 1965) to give estimates of the variance components. The fact that these explicit estimates exist suggests that iterative schemes based on first differentials might be available that give estimates for balanced designs in one iteration.

Animal breeding data that we deal with is seldom exactly balanced because of biological variation, for instance in litter size and sex of offspring, but experimenters usually have a target balanced design in mind. Under these circumstances one would hope a scheme which is optimal for balanced designs, i.e. yield estimates in one round of

iteration, would also be more efficient for unbalanced designs, i.e. converge faster than existing schemes. This idea is investigated for several designs and parameter values appropriate for animal breeding work.

2. MODELS WITH TWO VARIANCE COMPONENTS

In this section we consider model (1.1) with $p=1$ and $c=1$, including a between group component σ_1^2 and a within component σ^2 for several designs of increasing complexity.

a) Balanced one-way classification

The model can be written as:

$$y_{ij}=\mu+b_{1i}+e_{ij} \quad (i=1,\ldots,s, j=1,\ldots,m) \tag{2.1}$$

or in the formulation of (1.1) $p=1$ and $\mathbf{X}$ is a $sm\times 1$ matrix of ones, $\mu=\boldsymbol{\alpha}$, $q_1=s$ and $sm=n$ and $\mathbf{b}_1$ and $\mathbf{e}$ have elements b_i and e_{ij}. The log-likelihood of error contrasts (L) is given by

$$2L=-s(m-1)\ln\sigma^2-(s-1)\ln(\sigma^2+m\sigma_1^2)-W/\sigma^2-B/(\sigma^2+m\sigma_1^2) \tag{2.2}$$

where W and B are the within and between group sum of squares. The EM algorithm is based on treating μ, b_i and e_{ij} as "missing data" and arguing that if e_{ij} and b_i were known then one can form sufficient statistics

$$t=\Sigma e_{ij}^2 \quad \text{and} \quad t_1=\Sigma b_i^2 \tag{2.3}$$

and estimate σ^2 and σ_1^2 from

$$sm\sigma^2=t \quad \text{and} \quad s\sigma_1^2=t_1. \tag{2.4}$$

Because t and t_1 are not known Dempster *et al.* (1977) suggest that t and t_1 in (2.3) be replaced by their posterior expectation given σ_1^2 and σ^2 (an E or expectation step). As (2.4) gives a maximum

likelihood estimator this can be thought of as an M (or maximization step). Hence the two steps combine to give an EM algorithm. This gives rise to an iterative scheme increasing L in each iteration. An alternative, more direct, derivation is to differentiate L with respect to the "natural" parameters $-1/2\sigma^2$ and $-1/2\sigma_1^2$ and set the derivatives to zero (Dempster *et al.*, 1984). For model (2.1) it can be shown that one step of the EM algorithm changes σ^2 and σ_1^2 to $\hat{\sigma}^2$ and $\hat{\sigma}_1^2$ where

$$sm\hat{\sigma}^2 = W + \sigma^4 B^* + s\sigma^2 \tag{2.5}$$

and

$$s\hat{\sigma}_1^2 = s\sigma_1^2 + \sigma_1^4 m B^* \tag{2.6}$$

where

$$B^* = [B/(\sigma^2 + m\sigma_1^2) - (s-1)]/(\sigma^2 + m\sigma_1^2).$$

Table I gives an analysis of variance for such data. This gives estimates of σ^2 and σ_1^2 from

$$s(m-1)\sigma^2 = W \quad \text{and} \quad (s-1)(\sigma^2 + m\sigma_1^2) = B \tag{2.7}$$

which is of the same form as (2.4) with degrees of freedom, $s(m-1)$ and $(s-1)$, replacing the number of effects, sm and s. Further $\sigma^2 + m\sigma_1^2$ in (2.7) replaces σ_1^2 in (2.4). This suggests using σ^2, the residual variance, and $\theta_{1m} = \sigma_1^2 + \sigma^2/m$, the variance of a group mean, as an alternative parameterization, or more generally σ^2 and $\theta_{1k} = \sigma_1^2 + \sigma^2/k$, noting that as $k \to \infty$, $\theta_{1k} \to \sigma_1^2$.

TABLE I

Analysis of variance.

Source of variation	d.f.	Sum of squares	Expected mean square
Between groups	$s-1$	B	$\sigma^2 + m\sigma_1^2$
Within groups	$s(m-1)$	W	σ^2

Differentiating (2.1) with respect to $-1/2\sigma^2$ and $-1/2\theta_{1k}$ gives estimating equations:

$$s(m-1)\sigma^2 = W + (1-m/k)\sigma^4 B^* \tag{2.8}$$

$$(s-1)\theta_{1k} = (s-1)\theta_{ik} + \theta_{1k}^2 mB^*. \tag{2.9}$$

These equations reduce to (2.7) when $k=m$ and are a minor rearrangement of (2.5) and (2.6) as $k \to \infty$.

For a design with $m=25$ and $s=20$, Figures 1 and 2 show estimates of σ_1^2 found using (2.8) and (2.9) for $k=\infty$ and $k=25$, respectively. B and W were derived using $\sigma_1^2=0.125$ and $\sigma^2=0.875$ in Figure 1, and $\sigma_1^2=0.0125$ and $\sigma^2=0.9875$ in Figure 2. Whilst values of $\sigma_1^2/(\sigma^2+\sigma_1^2)$ are small, these are in the range of values often found in animal breeding studies. In both cases starting values of $\sigma_1^2=0.0625$ and $\sigma^2=0.9375$ were used. These figures emphasize the slow rate of convergence of the *EM* algorithm using the original parameters σ^2 and σ_1^2 ($k=\infty$) especially when σ_1^2/σ^2 is small.

b) Unbalanced classification

In order to investigate the effect of lack of balance, suppose we have p supergroups each with a different mean α_i and that within the ith supergroup we have a balanced structure of s_i groups, each of m_i observations ($i=1,\ldots,p$). Then if B_{1i} is the sum of squares between groups within supergroup i and W is the pooled within group sum of squares, estimates can be derived analogous to (2.8) and (2.9) based on σ^2 and θ_{1k} from

$$[\Sigma s_i(m_i-1)]\hat{\sigma}^2 = W + \sigma^4[\Sigma[(1-m_i/k)B_i^*] \tag{2.10}$$

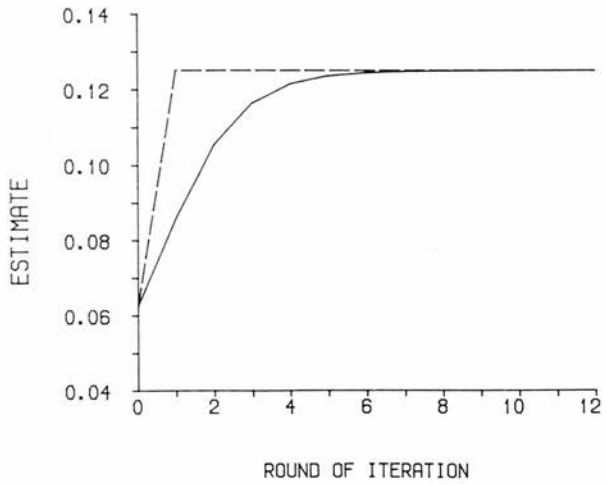
and

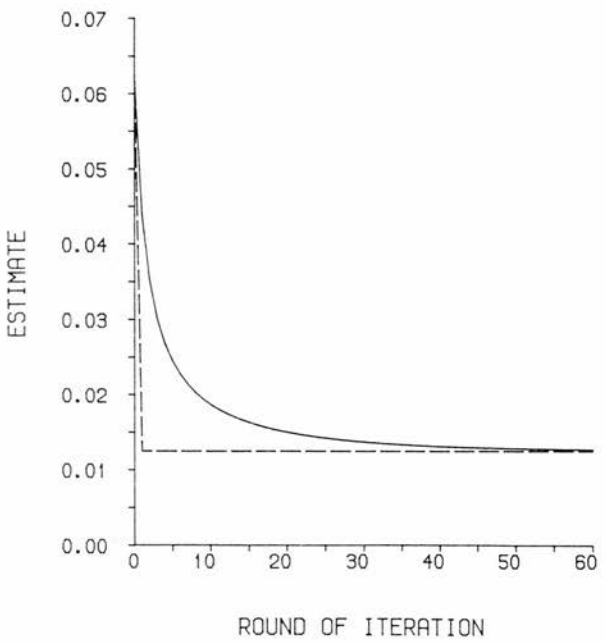
$$[\Sigma(s_i-1)]\hat{\theta}_{1k} = [\Sigma(s_i-1)]\theta_{ik} + \theta_{ik}^2[\Sigma m_i B_i^*] \tag{2.11}$$

where

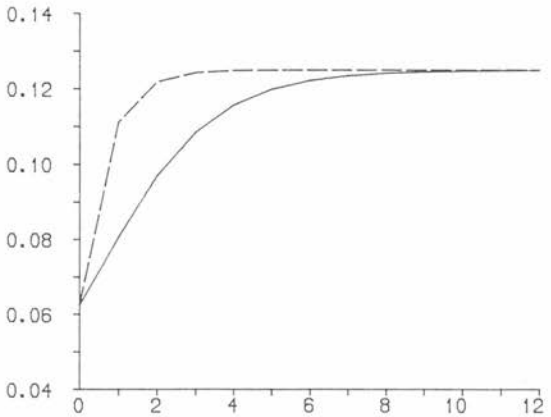
$$B_i^* = [B_i/(\sigma^2+m_i\sigma_1^2) - (s_i-1)]/(\sigma^2+m_i\sigma_1^2).$$

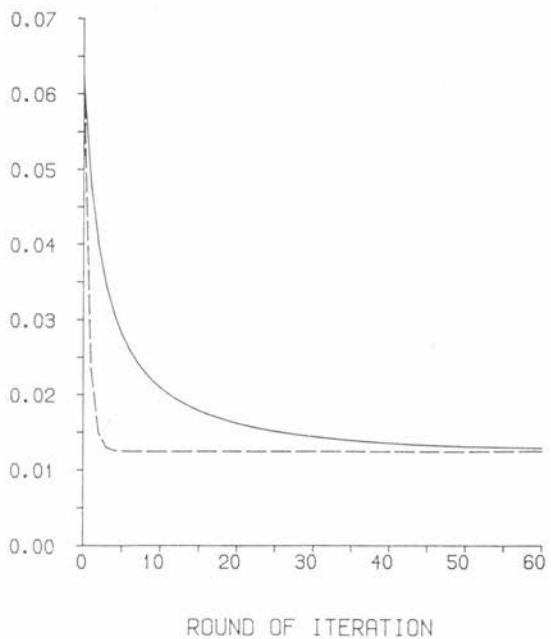
Figures 3 and 4 show estimates of σ_1^2 over rounds of iteration obtained using (2.10) and (2.11). The design examined had $p=5$





Figures 1 and 2 Values of σ_1^2 from an iterative scheme for a balanced design $k=\infty$ =solid curve; $k=25$ =dotted curve.





FIGURES 3 and 4 Values of σ_1^2 from an iterative scheme for an unbalanced design. $k=\infty$ = solid curve, $k=25$ = dotted curve.

supergroups with four groups in each supergroup. The respective group sizes were $m_i = 15$, 20, 25, 30 and 35 with a mean of 25. Starting values and values of σ^2 and σ_1^2 used to derive B_i and W were as for Figures 1 and 2. Again the untransformed EM algorithm ($k = \infty$) is contrasted to the reparameterization with $k = 25$, the mean family size. For the latter estimates of σ_1^2 converge quickly, whilst the rate of convergence for $k = \infty$ is similar to that for the balanced design (see Figures 1 and 2).

c) Mixed model

Slightly more complicated formulae are needed if the model of analysis (1.1) includes estimable fixed effects $p(>1)$.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}_1\mathbf{b}_1 + \mathbf{e}.$$

Then weighted least squares estimates of α can be found from

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_1 \\ \mathbf{Z}_1'\mathbf{X} & \mathbf{Z}_1'\mathbf{Z}_1 + \mathbf{I}\gamma \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}_1'\mathbf{y} \end{bmatrix}.$$

With $\gamma = \sigma^2/\sigma_1^2$, or rearranging (2.11)

$$\begin{bmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xz} \\ \mathbf{C}_{zx} & \mathbf{C}_{zz} \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}_1'\mathbf{y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}.$$

It can be shown by manipulating formulae given by Harville (1977) and Patterson and Thompson (1971), that an iterative scheme analogous to (2.5) and (2.6) is given by

$$(n-p)\sigma^2 = \mathbf{e}'\mathbf{e} - \boldsymbol{\beta}_1'\boldsymbol{\beta}_1\sigma^4/k\sigma_1^4 + q\sigma^2 + q\sigma^4/k\sigma_1^2 - tr(\mathbf{C}_{zz})\theta_{1k}\sigma^4/\sigma_1^4, \tag{2.12}$$

$$q\hat{\theta}_{1k} = q(\theta_{ik} - \theta_{1k}^2/\sigma_1^2) + (\theta_{1k}/\sigma_1^2)^2[\boldsymbol{\beta}_1'\boldsymbol{\beta}_1 + \sigma^2 tr(\mathbf{C}_{zz})], \tag{2.13}$$

where $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{Z}\boldsymbol{\beta}$, and q, as defined in (1.1), is the number of levels of the random effect.

This method was applied to a reproductive study recently considered by Dempster *et al.* (1984). Estimates of σ_1^2 and σ^2 for different choices of k are given in Table II showing the improvement in reducing k from infinity.

TABLE II

Iterative values of parameters for reproductive data.

	Estimates of σ_1^2			Estimates of σ^2		
Iteration	$k=\infty$	$k=11.93$	$k=9$	$k=\infty$	$k=11.93$	$k=9$
0	0.064289	0.064289	0.064289	0.251014	0.257014	0.251014
1	0.076860	0.093968	0.100165	0.162718	0.161212	0.160725
2	0.084268	0.093814	0.097581	0.163220	0.162825	0.162847
3	0.094367	0.097305	0.097419	0.162922	0.162801	0.162802
4	0.096299	0.097384	0.097402	0.162843	0.162802	
5	0.097004	0.097397	0.097400	0.162816		
6	0.097258	0.097399		0.162807		
7	0.097344	0.097400		0.152803		
8	0.097381			0.162802		
9	0.097393					
10	0.097397					

Patterson and Thompson (1971) and Dempster *et al.* (1984) note that the matrix inversion needed to find $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_1$ in (2.11) for each round of (2.12) and (2.13) can be avoided if a spectral decomposition of $\mathbf{Z}_1'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Z}=\mathbf{PDP}'$ is used. The matrix $\mathbf{P}$ is an orthogonal matrix $\mathbf{D}$ is a diagonal matrix with elements λ_i of which r are non-zero elements. This can be used to construct an analysis of variance with r sum of squares b_i each with one degree of freedom, formed from the squares of the vector $\mathbf{P}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}$ with expectation $\sigma^2+\lambda_i\sigma_1^2$. There is also a within group of squares $W=\mathbf{e}'\mathbf{e}$ found using $\gamma=0$ to evaluate $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_1$. This analysis of variance is similar to that for the unbalanced classification (Section 2b) and so (2.9) and (2.10) could be used replacing the residual degrees of freedom $\Sigma s_i(m_i-1)$ in (2.9) by n-rank $[\mathbf{XZ}_1]$ and replacing (s_i-1) by 1 and m_i by λ_i.

3. HIERARCHICAL DESIGN WITH THREE VARIANCE COMPONENTS

A common animal breeding design is an hierarchical design when each father is mated to several mothers and several offspring are raised and measured from each mating. There is interest in estimating the father (σ_2^2), mother (σ_1^2) and residual (σ^2) variance components, i.e. $c=2$ in model (1.1). Suppose there are p supergroups each with a balanced structure. In the ith supergroup observations have mean subscript i and there are measurements on $s_i d_i m_i$ animals. Each of s_i fathers is mated to d_i mothers and m_i offspring from each mating are measured. If the design was balanced with $s_i=s$, $d_i=d$ and $m_i=m$ for $i=1,\ldots,p$, a hierarchical analysis of variance could be constructed and the mean squares between fathers (F), between mothers within fathers (M) and within mothers (W) could be used to estimate $\sigma^2+d\sigma_1^2+md\sigma_2^2$, $\sigma^2+m\sigma_1^2$ and σ^2, respectively. The argument of Section 2 suggests for this unbalanced design using $\theta_{1k_1}=\sigma_1^2+\sigma^2/k_1$ and $\theta_{2k_2}=\sigma_2^2+(\theta_{1k_1})/k_2=\sigma_2^2+\sigma_1^2/k_2+\sigma^2/k_1k_2$ as parameters.

An iterative solution schemes based on σ^2, θ_{1k_1} and θ_{2k_2} is given by

$$[s_i d_i(m_i-1)]\hat{\sigma}^2 = W+\sigma^4\Sigma(1-m_i/k_1)[M_i^*+F_i^*] \tag{3.1}$$

$$[\Sigma s_i(d_i-1)]\hat{\theta}_{1k_1}=[\Sigma s_i(d_i-1)]\theta_{1k_1}+\theta_{1k_1}^2\Sigma[m_iM_i^*+(1-d_i/k_2)m_iF_i^*] \tag{3.2}$$

$$[\Sigma(s_i-1)]\hat{\theta}_{2k_2}=[\Sigma(s_i-1)]\theta_{2k_2}+\theta_{2k_2}^2\Sigma[m_id_iF_i^*] \tag{3.3}$$

where

$$F_i^*=(F_i/(\sigma^2+m_i\sigma_1^2+m_id_i\sigma_2^2)-(s_i-1))/(\sigma^2+m_i\sigma_1^2+m_id_i\sigma_2^2)$$

$$M_i^*=(M_i/(\sigma^2+m_i\sigma_1^2)-s_i(d_i-1))/(\sigma^2+m_i\sigma_1^2).$$

F_i denotes the sums of squares between fathers, and M_i the sum of squares between mothers within fathers in the ith group, and W is the within mother sum of squares. For a balanced design $m_i=m$, $d_i=d$ and $s_i=s$. Then (3.1), (3.2) and (3.3) with $k_1=m$ and $k_2=d$ reduce to $psd(m-1)\sigma^2=W$, $ps(d-1)m\theta_{1m}=\Sigma M_i$, $p(s-1)m\theta_{2d}=\Sigma F_i$, i.e. as for the balanced one-way classification estimates converge in one round of iteration.

A design was constructed with 12 supergroups and $s_i = 5$ for each supergroup. To generate m_i and d_i, four values of d_i, 2, 4, 6 and 8, and three values of m_i, 3, 5 and 7, were chosen and each combination of d_i and m_i was used in one supergroup. Figures 5 and 6 show estimates for σ_1^2 (the slowest converging parameter) using $k_1 = k_2 = \infty$ and $k_1 = k_2 = 5$ in (3.1), (3.2) and (3.3).

W, F_i and M_i were derived using $\sigma^2 = 0.75$ $\sigma_1^2 = \sigma_2^2 = 0.125$ (Figure 5) and $\sigma^2 = 0.975$ $\sigma_1^2 = \sigma_2^2 = 0.0125$ (Figure 6) and starting values were $\sigma^2 = 0.875$ and $\sigma_1^2 = \sigma_2^2 = 0.0625$. Again the figures show the advantage of reducing k_1 and k_2 from infinity.

4. MULTIVARIATE VARIANCE COMPONENTS

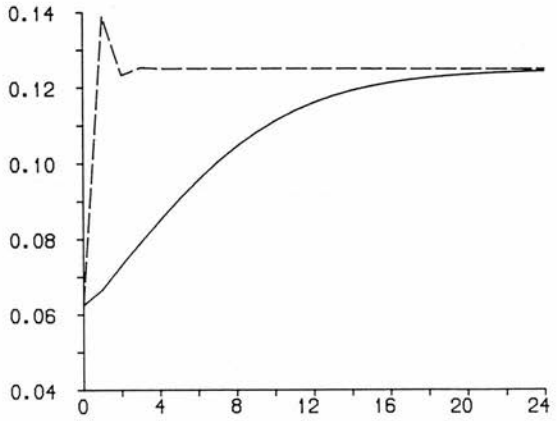
In this section we consider a multivariate generalization motivated by a problem considered by Meyer (1983). We suppose t variates can be measured and that a model of the form (1.1) is appropriate for each variate with $c = 1$. We assume that the random effects for the ith and jth traits have covariance $\sigma_{1ij} I$ and the errors terms have covariance $\sigma_{ij}\mathbf{I}$. There are then $t(t+1)$ variance parameters σ_{1ij} and σ_{ij} $(j \leqq 1,\ i = 1, \ldots, t)$ which can be written as two $t \times t$ symmetric matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_1$. One simple design, a generalization of Section 2b, is to assume that observations are in p supergroups and that in the ith supergroup t_i variates are measured on m_i individuals in s_i groups. The mean effect for the jth variate in the ith group is α_{ij}. For example with $p = 2$ supergroups and $t = 2$ variates measured, we might have the first variate measured on m individuals in each of s_i groups in the first supergroup and both variates measured on m_2 individuals in each of s_2 groups in the second supergroup.

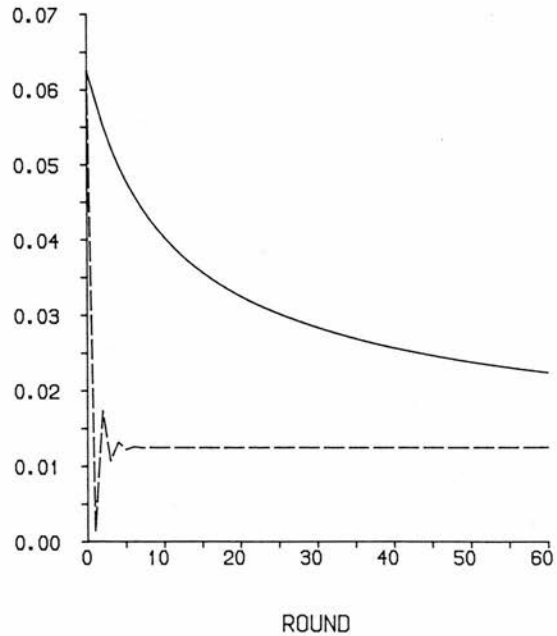
The log-likelihood of error contrasts for the ith supergroup (L_i) is given by

$$2L_i = -(s_i - 1)\ln\left|\boldsymbol{\Sigma}_i + m_i\boldsymbol{\Sigma}_{1i}\right| - s_i(m_i - 1)\ln\left|\boldsymbol{\Sigma}_i\right| - \operatorname{trace}(\boldsymbol{\Sigma}_i^{-1}\mathbf{W}_i)$$

$$- \operatorname{trace}\left((\boldsymbol{\Sigma}_i + m_i\boldsymbol{\Sigma}_{1i})^{-1}\mathbf{B}_i\right)$$

where $\mathbf{W}_i$ and $\mathbf{B}_i$ are the within and between sum of squares and cross-products for the t_i variates measured in ith group, and $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Sigma}_{1i}$ are the subset of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_1$ relevant to the measurements in the





FIGURES 5 and 6 Values of σ_1^2 from an iterative scheme for an unbalanced hierarchical design, $k_1 = k_2 \infty$ = solid curve; $k_1 = k_2 = 5$ dotted curve.

ith supergroup. The argument in the previous section suggests using $\boldsymbol{\Sigma}$ and $\boldsymbol{\theta}_{1k}$ instead of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_1$ as parameters where $\boldsymbol{\theta}_{1k} = \mathbf{K} \odot \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}$ and the matrix $\mathbf{K}$ is a symmetrical matrix with elements K_{ij} $(j \leqq i,\ i = 1, \ldots, t)$ and $\odot$ is an operator that multiplies two matrices term by term, i.e. the ijth element of $\mathbf{A} \odot \mathbf{B}$ has (i, j) element $A_{ij}B_{ij}$.

An iterative scheme for $\boldsymbol{\Sigma}$ and $\boldsymbol{\theta}_{1k}$ is then

$$\mathbf{D} \odot \boldsymbol{\Sigma} = D \odot \boldsymbol{\Sigma} + \boldsymbol{\Sigma}[dL/d\boldsymbol{\Sigma}]\boldsymbol{\Sigma} \tag{4.1}$$

$$\mathbf{D}_1 \odot \boldsymbol{\theta}_{1k} = \mathbf{D}_1 \odot \boldsymbol{\theta}_{1k} + \boldsymbol{\theta}_{1k}[dL/\boldsymbol{\theta}_{1k}]\boldsymbol{\theta}_{1k} \tag{4.2}$$

where $dL/d\boldsymbol{\phi}$ $(\boldsymbol{\phi} = \boldsymbol{\Sigma}, \boldsymbol{\theta}_{1k})$ is a symmetric matrix defined by diagonal elements $dL/d\phi_{ii}$ and off diagonal elements $(\frac{1}{2})\ dL/d\phi_{ij}$ $(i < j)$. Correspondingly, $\mathbf{D}$ and $\mathbf{D}_1$ are symmetric matrices with (i, j)th element equal to the degrees of freedom within and between groups, for the ith and jth trait respectively.

For the example above with two supergroups then

$$\mathbf{D} = \begin{bmatrix} s_1(m_1 - 1) + s_2(m_2 - 1) & s_2(m_2 - 1) \\ s_2(m_2 - 1) & s_2(m_2 - 1) \end{bmatrix} \quad \text{and}$$

$$\mathbf{D}_1 = \begin{bmatrix} s_1 - 1 + s_2 - 1 & s_2 - 1 \\ s_2 - 1 & s_2 - 1 \end{bmatrix}.$$

As a multivariate example, we take a design with

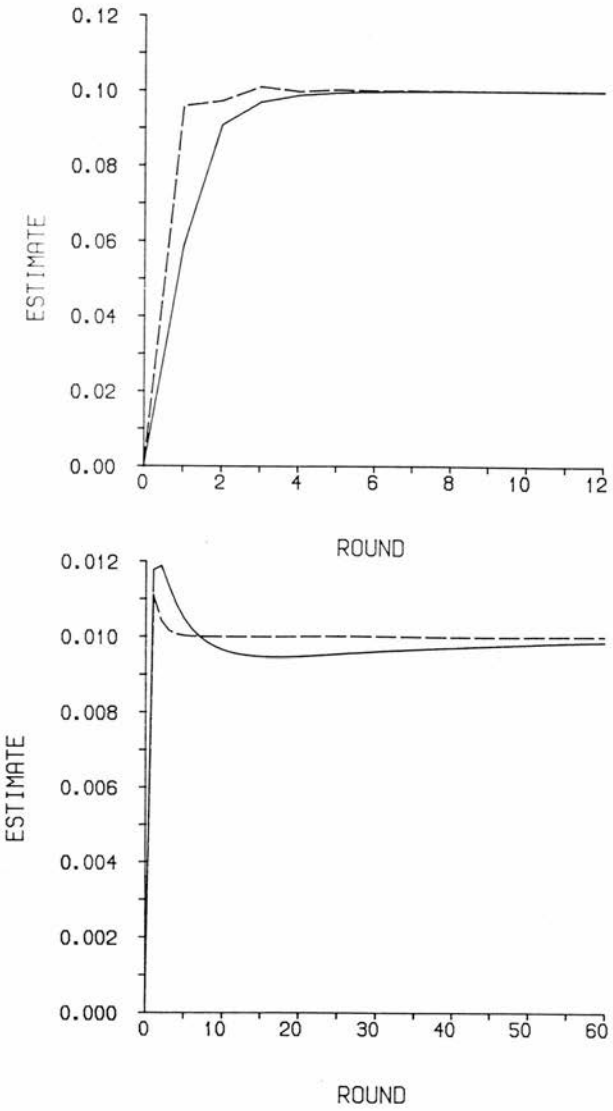
$$s_1 = 50,\ s_2 = 50,\ m_1 = 10,\ m_2 = 40$$

and calculate W_i and B_i using

i) $$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.125 & 0.100 \\ 0.100 & 0.125 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.875 & 0.400 \\ 0.400 & 0.875 \end{bmatrix}$$

and

ii) $$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.0125 & 0.0100 \\ 0.0100 & 0.0125 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.9875 & 0.4900 \\ 0.4900 & 0.9875 \end{bmatrix}$$



FIGURES 7 and 8 Values of σ_{112} from an iterative scheme for an unbalanced multivariate design.

$$k=\begin{bmatrix}\infty & \infty\\ \infty & \infty\end{bmatrix}=\text{solid curve;}\qquad k=\begin{bmatrix}50 & 40\\ 40 & 40\end{bmatrix}=\text{dotted curve.}$$

and starting values

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.0625 & 0.0000 \\ 0.0000 & 0.0625 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.9375 & 0.0000 \\ 0.0000 & 0.9375 \end{bmatrix}.$$

These estimates and parameter values are a multivariate extension of those used in Section 2. Figures 7 and 8 show graphs of estimates of σ_{12}, using (4.1) and (4.2) for cases (i) and (ii), using values for

$$\mathbf{K} = \begin{bmatrix} \infty & \infty \\ \infty & \infty \end{bmatrix} \quad \text{and} \quad \mathbf{K} = \begin{bmatrix} 50 & 40 \\ 40 & 40 \end{bmatrix}.$$

5. CONCLUSION

It has been shown that in a variety of settings, consideration of balanced designs suggest a reparameterization that improves convergence. These schemes have been constructed directly from the first differentials of L but, no doubt, could have been derived by considering linear functions, of $\boldsymbol{\alpha}$, $\mathbf{b}_1$ and $\mathbf{e}$, such as $\mathbf{b}_1 - (1/k_1)\mathbf{Z}_1'\mathbf{Z}_1\mathbf{e}$ and treat them as "missing values" in the *EM* algorithm. There are other variance component problems with non-normal data (Laird (1978), Harville and Mee (1984)) where this approach is probably useful.

One property of the *EM* algorithm is that the variance component estimates will always be greater than or equal to zero (Harville (1977)). By contrast the scheme outlined can lead to negative estimates. However the *EM* algorithm becomes slower to converge as variance components tend to zero and we have found it more practical to use our scheme and then constrain or attenuate the parameter estimates if we wish to impose restrictions on the estimates.

References

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood with incomplete data via the *EM* algorithm. *Journal of the Royal Statistical Society Series B*, **39**, 1–38.

Dempster, A. P., Selwyn, M. R., Patel, C. M. and Roth, A. J. (1984). Statistical and computational aspects of mixed model analysis. *Applied Statistics*, **33**, 203–214.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72,** 320–340.

Harville, D. A. and Mee, R. W. (1984). A mixed-model procedure for analysing ordered categorial data. *Biometrics*, **40,** 393–408.

Laird, N. M. (1978). Empirical Bayes methods for two-way contingency tables. *Biometrika*, **65,** 581–590.

Laird, N. M. (1982). Computation of variance components using the *EM* algorithm. *J. Statist. Comput. Simul.*, **14,** 295–303.

Meyer, K. (1983). Maximum likelihood procedures for estimating genetic parameters for later lactations of dairy cattle. *J. Dairy Science*, **66,** 1988–1997.

Nelder, J. A. (1965). The analysis of randomized experiments with orthogonal block structure: Block structure and the null analysis of variance. *Proc. Roy. Soc. A.*, **283,** 147–162.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58,** 545–554.

Robinson, D. L., Digby, P. G. N. and Thompson, R. (1982). *REML* a program for the analysis of non-orthogonal data by restricted maximum likelihood. *COMPSTAT 82, Part II*, Physica-Verlag, Vienna.

Thompson, R. (1977a). Discussion of the paper by Dempster *et al. Journal of Royal Statistical Society*, **39,** 34.

Thompson, R. (1977b). The estimation of heritability with unbalanced data. *Biometrics*, **33,** 485–495.

Thompson, R. (1979). Sire evaluation. *Biometrics*, **35,** 339–354.

11

A program for the analysis of non-orthogonal data by restricted maximum likelihood

by

D.L. Robinson, R. Thompson and P.G.N. Digby

REML - A Program for the Analysis of Non-orthogonal data by Restricted Maximum Likelihood.

D L Robinson, R Thompson and P G N Digby.

In any unbalanced experiment with more than one factor, good estimates of treatments must take into account the lack of balance in the levels of the other factors. With fixed effects, this may be achieved by regression. If there are also unbalanced blocking factors which are not orthogonal to treatment factors, for example in incomplete block designs, then efficient treatment estimates can only be found when the variances of these blocking factors are known, enabling recovery of inter-block information.

An appropriate analytical technique is that of Restricted Maximum Likelihood (Patterson and Thompson, 1971) which gives unbiased estimates identical to those from an analysis of variance in the case of balanced data. As well as the efficient estimation of fixed effects. The magnitude of sources of variation can be identified and effects related to the random factors can be predicted.

These are important; for example in sire evaluation it is usual to assume that a daughter's yield had a fixed herd-year-season effect and a random sire effect. From this inherently unbalanced dataset prediction of the random sire effect is required, so that chosen sires can later be used extensively, on the basis of their daughter's yields. The magnitude and sources of variation are also important so that testing strategies can be evaluated, and optimal schemes used.

Such an analysis has been programmed in standard Fortran, based on a modification of the algorithm proposed by Thompson (1977). The usual regression equations are reformulated as $(X'X+G^{-1})\underline{a}=X'\underline{y}$, where X is obtained from the design matrix for fixed and random effects, and $\underline{y}$ is a vector of observed values. Random effects are incorporated by the diagonal matrix G^{-1} , whose ith entry is 0 if the ith parameter a_i in $\underline{a}$ is fixed, and $1/v$ if a_i is a random effect from a population with variance v . An iterative technique, similar to Fisher's scoring technique, is used to solve the likelihood equations. Initial values of variance components are formed, or supplied by the user, and used to estimate the parameter vector $\underline{x}$; new variance components are then found from the derivatives of the likelihood equations.

Since X'X can be written in a form that is largely block diagonal, most of the parameter estimates can be formed sequentially using small submatrices which occupy the same storage locations, only the relevant parameters being retained.

The estimates for the likelihood equations are accumulated at the same time. Computing time is saved since only small matrices need to be inverted, and storage requirements are lessened considerably.

A wide range of linear models may be fitted, with no restriction on the number of fixed or random factors or covariates. Input is by simple directives specifying the factors, variates and covariates in the model to be fitted. Storage is allocated dynamically, so the directives may appear in almost any order. The data can be read sequentially or in parallel, or any combination of the two, and in free or fixed format.

The user can specify the amount of output required, from simple, easy to read, tables of selected estimates of effects, optionally with average standard errors, to a complete list of all the parameter estimates and tables of pairwise standard errors at every iteration. The former will be useful if there are factors with a large number of levels which are not of interest and would clutter the output, the latter may aid interpretation of the data. Often the most useful combination is to print (i) the components of variance evey iteration or two to check on convergence, and (ii) the required estimates of effects and variances after the final iteration. Approximate stratum variances based on a canonical decomposition of the information matrix are also available. In addition a full program monitor can be used to check on some or all of the calculations performed.

Jennrich and Sampson (1978) ask "What if anything can be done about the fact that general purpose maximum likelihood programs use a lot of computer storage placing rather severe size limitations on the size of problems that can be handled?". By the sequential formation of small submatrices occupying the same storage locations REML overcomes this problem and has fitted models to datasets with over 9,000 observations, some factors having ~~nearly~~ 1,400 levels.

References.

Jennrich, R I, and Sampson, P F (1978). Some problems faced in making a variance component algorithm into a general mixed model program. Proceedings of the 11th annual symposium on the interface between computer science and statistics, Raleigh, 1978, 56-63.

Patterson, H D, and Thompson, R (1971). Recovery of inter-block information when block sizes are unequal. Biometrics 58, 445-554.

Thompson, R (1977). The estimation of heritability with unbalanced data. Biometrics 33, 485-495.

12

Diallel crosses, partially balanced inccomplete block designs with triangular association schemes and rectangular lattices

by

R. Thompson

DIALLEL CROSSES, PARTIALLY BALANCED INCOMPLETE BLOCK DESIGNS WITH TRIANGULAR ASSOCIATION SCHEMES AND RECTANGULAR LATTICES

INTRODUCTION

What have these designs in common? Apart from esoteric names, all three have a high degree of symmetry and yet cannot be completely analysed by the Genstat 'ANOVA' directive. The treatment structure for the first design cannot be specified, the 'ANOVA' algorithm finds the second design unbalanced and treatment means for the third are only 'approximate'. Of course, other methods of analysis are possible and macros are available to analyse (i) diallel crosses using regression procedures (N.G. Alvey and O. Mayo, personal communications) and (ii) rectangular lattices using 'ANOVA' on the dual of the design (E.R. Williams and D. Ratcliff, CSIRO Macro).

I suggest instead a technique which aids in the analysis and, I think, the understanding of all three designs (and no doubt others). It is based on the idea of repeating the data several times and generating treatment and block factors with each copy of the data. The treatment and block factors are chosen so that the repeated data has a 'balanced' structure and that the required analysis is a meaningful subset (usually the only one) of the analysis of the extended data. Thompson and Baker (1981) found a similar idea useful in embedding several types of data into a generalised linear model framework.

I will demonstrate the technique for each design in turn and then discuss the disadvantages of the method and point out how the Genstat 'ANOVA' algorithm might be improved.

DIALLEL CROSSES

In plant breeding work on p different parental lines, often, all reciprocal crosses can be made and the p^2 progeny used for the comparison of parental lines. Various models and analyses have been suggested (Yates (1947), Hayman (1954), Griffing (1956), Jinks and Mather (1971)). I first consider a simple model that exhibits the difficulty in fitting these classes of models. I take the case where the progeny mean, when a male of line i is crossed with a female of line j, is y_{ij}. One might think of fitting a factorial model

$$y_{ij} = \textit{mean + male line (i) effect + female line (j) effect}$$

and this can be easily done in 'ANOVA'. However it might be reasonable to assume the male and female effects of line i to be similar and so want to fit a model

$$y_{ij} = \textit{mean + line (i) effect + line (j) effect.} \qquad (1)$$

However, each observation has 2 contributions from the line factor and so it is difficult to specify the treatment structure. Suppose, however, that we have two observations yt_{ij1} and yt_{ij2} with expectations

$$yt_{ij1} = \textit{mean + line (i) effect}\ , \quad yt_{ij2} = \textit{mean + line (j) effect.}$$

Each such pair yt_{ij1} and yt_{ij2} can be put in a different level of a factor DUMMY (for example put yt_{ij1} and yt_{ij2} into level $(i+(j-1)xp)$. For $p=3$ the data, treatment and block factors will be

$$\begin{aligned}
Y &= (y_{11} \quad y_{12} \quad y_{13} \quad y_{21} \quad y_{22} \quad y_{23} \quad y_{31} \quad y_{32} \quad y_{33}\ ) \\
YT &= (yt_{111} \quad yt_{121} \quad yt_{131} \quad yt_{211} \quad yt_{221} \quad yt_{231} \quad yt_{311} \quad yt_{321} \quad yt_{331} \\
&\quad\ yt_{112} \quad yt_{122} \quad yt_{132} \quad yt_{212} \quad yt_{222} \quad yt_{232} \quad yt_{312} \quad yt_{322} \quad yt_{332}) \\
&= (y_{11} \quad y_{12} \quad y_{13} \quad y_{21} \quad y_{22} \quad y_{23} \quad y_{31} \quad y_{32} \quad y_{33} \\
&\quad\ y_{11} \quad y_{12} \quad y_{13} \quad y_{21} \quad y_{22} \quad y_{23} \quad y_{31} \quad y_{32} \quad y_{33}\ ) \\
LINE &= (1 \quad 1 \quad 1 \quad 2 \quad 2 \quad 2 \quad 3 \quad 3 \quad 3 \\
&\quad\ 1 \quad 2 \quad 3 \quad 1 \quad 2 \quad 3 \quad 1 \quad 2 \quad 3\ ) \\
DUMMY &= (1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \\
&\quad\ 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9\ ).
\end{aligned}$$

Then an analysis of the p^2 x 2 values yt_{ijk} with LINE as treatment factor and DUMMY as a blocking factor will use *1* in the DUMMY stratum and, hence, give the required analysis in the DUMMY stratum. As noted by Wilkinson (1971) the resulting line effects have efficiency $(p-2)/(p-1)$.

Other more complicated models might be fitted. In Appendix 1 the code is given for Hayman's (1954) analysis using data in Mather and Jinks (1971, p. 256). This analysis includes comparisons of mean parental performance with mean progeny performance (i.e. y_{ii} versus y_{ij} $(j=i)$ or 'pure' versus 'cross') and of line parental performance with line cross performance and fits a model similar to (1) to the reciprocal difference $(y_{ij}-y_{ji})$.

PARTIALLY BALANCED INCOMPLETE BLOCKS WITH TRIANGULAR ASSOCIATION SCHEMES (P.B.I.B.T.)

These designs (for example Clatworthy, 1973) have $t = n(n-1)/2$ treatments. The association scheme can be obtained from a symmetric n x n array A with the diagonal blank and each treatment occurring twice - once above and once below the diagonal. Treatments in the same row or column are first associates. For example, the design with $t=10$, $b=6$, $k=5$, $r=3$, with treatments

1, 3, 7, 8, 9	in the first block
1, 2, 6, 9, 10	in the second block
1, 4, 5, 8, 10	in the third block
2, 4, 6, 7, 8	in the fourth block
2, 3, 5, 7, 10	in the fifth block and
3, 4, 5, 6, 9	in the sixth block

is a P.B.I.B.T..

The design has array *A*

*	*1*	*2*	*3*	*4*
1	*	*5*	*6*	*7*
2	*5*	*	*8*	*9*
3	*6*	*8*	*	*10*
4	*7*	*9*	*10*	*

and this indicates that *2, 3, 4, 5, 6* and *7* are the first associates of *1* and *3, 6, 8, 4, 7* and *9* are first associates of *10*. As *A* is symmetric, the information on the association scheme can be found from the triangular array above (or below) the diagonal and hence the designs are said to have a triangular association scheme. The designs cannot be analysed directly by 'ANOVA' but they can be made balanced by imposing a pseudo-treatment structure. Because of the association scheme it is natural to use the rows

and columns of A to generate some pseudo-treatment structure. Suppose that treatment i (in the lower triangle of A) is in row r_i and column c_i. Then T.P. Speed has shown that P.B.I.B.T. designs have general balance with respect to the pseudo-treatment scheme

$$i\text{-th treatment effect} = r_i\text{-th pseudo-treatment effect} + c_i\text{-th pseudo-treatment effect} + r_i, c_i \text{ pseudo-treatment interaction effect}$$

with $(n-1)$ degrees of freedom for the pseudo-treatment effects and $(n-2)(n-1)/2$ degrees of freedom for the interaction effects.

This model is similar to the diallel model and can be fitted in a similar way. We make 2 copies of the data and use the rows of A to give the levels of the pseudofactor for the first copy and the columns of A to give the levels of the pseudofactor for the second. For example, the treatments

1 2 3 4 5 6 7 8 9 10

can be replaced by

2 3 4 5 3 4 5 4 5 5

with the first copy of the data and by

1 1 1 1 2 2 2 3 3 4

with the second copy of the data. Appendix 2 gives the Genstat code for the analysis of variance for this design, with some artificial data.

One problem is that Genstat does not know that each treatment is associated with 2 levels of the pseudofactor and it is not clear how the levels are calculated when deriving tables of means. For instance with the example above, Genstat associates levels *2, 2, 3, 3, 3, 3, 4, 4, 4, 5* with the treatments *1-10*. The level for treatment i is the integer part of $[(r_i+c_i+1)/2]$. In most other circumstances, the association of two levels of a pseudofactor with one treatment combination will only occur by mistake and it is intended to trap this in future releases of Genstat! However, correct tables of means and standard errors can easily be constructed using 'EXTRACT'.

A macro is available from the author to generate the levels of the pseudo-factors from treatment and block-factors for P.B.I.B.T. designs when A is not known, for example, if the treatments are randomised and no note is kept of the original order.

RECTANGULAR LATTICES

These designs, introduced by Harshbarger (1950) for $p(p-q)$ treatments in blocks of $(p-q)$, are a development of square lattices (Yates, 1936). Macros exist to analyse rectangular lattices but it is interesting to show how the method of the previous sections can also be used with these designs.

Consider a design for *12 = 4* x *3* treatments in 3 replications and block size of 3. A rectangular lattice design with these dimensions can be constructed from the following Graeco-Latin square

TP 1 \ TP 2	1	2	3	4
1	δ 13 A	α 1 D	β 2 B	γ 3 C
2	α 4 C	δ 14 B	γ 5 D	β 6 A
3	β 7 D	γ 8 A	δ 15 C	α 9 B
4	γ 10 B	β 11 C	α 12 A	δ 16 D

where the levels of *TP3* are the Latin letters *A*, *B*, *C*, *D* and of *TP4* are the Greek Letters α§ β§ γ§ δ. We associate with each cell a treatment, the off-diagonal cells are numbered from *1* to *12*, the diagonal from *13* to *16*. Then, using rows, columns and treatments to generate blocks, a 3-replicate rectangular lattice design can be derived by ignoring the diagonal elements of the square. (It will become apparent later why we have bothered to label the diagonal treatments as *13* to *16*.)

This is

Rep I				*Rep II*				*Rep III*			
	1	*2*	*3*		*4*	*7*	*10*		*6*	*8*	*12*
	4	*5*	*6*		*1*	*8*	*11*		*2*	*9*	*10*
	7	*8*	*9*		*2*	*5*	*12*		*3*	*4*	*11*
	10	*11*	*12*		*3*	*6*	*9*		*1*	*5*	*7*

This suggests using *TP1 + TP2 + TP3 + TP4* as a pseudofactor structure for *T* and, although this allows Genstat to produce the correct analysis of variance, the treatment effects are only 'approximate'. This is because *TP1*, *TP2* and *TP3* are not mutually orthogonal.

T.P. Speed has suggested an alternative subdivision of the pseudofactor space which allows an orthogonal subdivision. In our case, rather than think of *TP1*, *TP2*, *TP3* as representing 3 sets of 4 effects, we think of them as a two-way table indexed by *COPY* with 3 levels and *TP* with 4 levels, i.e.

		TP1	*TP2*	*TP3*
COPY		*1*	*2*	*3*
	1	*1*	*1*	*A*
TP	*2*	*2*	*2*	*B*
	3	*3*	*3*	*C*
	4	*4*	*4*	*D*

Speed has shown that *TP* and *TP.COPY* generate orthogonal subspaces (with 3 degrees of freedom and 6 degrees of freedom in our case). Note that the combinations of the pseudofactors *TP1, TP2, TP3* at level i of *TP* relate to the i-th diagonal of the Graeco-Latin square, so that the average effects of *TP* would give estimates to compare treatments *13* to *16* if the pseudo-factorial model was appropriate.

Hence, each treatment is now associated with 3 levels of *TP* and *TP.COPY*. As before, we can specify this model using 3 copies of the data and using *TPi* to give the level of *TP* with the ith copy of the data. Appendix 3 gives the Genstat code for a rectangular lattice taken from Kempthorne (1952, p.522). Again, there is a problem linking up treatment estimates with the 3 levels of TP and COPY.

DIFFICULTIES

There are three disadvantages of this method (i) multiple copies of, for instance, the yield variate are needed, (ii) effects and sums of squares in the dummy stratum are calculated, (iii) the linkage between treatments and pseudofactors is not complete.

G.N. Wilkinson has suggested that a simple modification of the algorithm should avoid at least the first two of the difficulties. He would allow the treatment factor length to be a multiple, say c, of the length of the analysed variate. When treatment totals are calculated, the program would go through the data c times and when sweeps are made, c terms would be taken from each data value. A similar technique could be used for deriving treatment estimates from pseudofactors.

REFERENCES

Clatworthy, W.H. (1973) Tables of two-associate-class partially balanced designs (US Dept of Commerce).

Griffing, B. (1956) A generalised treatment of the use of diallel crosses in quantative inheritance. Heredity, 10, 31-50.

Harshbarger, B. (1950) Triple rectangular lattices. Biometrics, 5, 1-13.

Hayman, B.I. (1954) The analysis of variance of diallel tables. Biometrics, 10, 235-244.

Jinks, J.L. and Mather, K. (1971) Biometric genetics. (Chapman and Hall).

Kempthorne, O. (1952) The design and analysis of experiments. J. Wiley and Sons, New York.

Thompson, R. and Baker, R.J. (1981) Composite link functions in generalised linear models. J.R.S.S., 30, 123-131.

Wilkinson, G.N. (1970) A general recursive algorithm for analysis of variance. Biometrika, 57, 19-46.

Yates, F. (1936) A new method of arranging variety trials involving a large number of varieties. Journal of Agricultural Science, 26, 424-455.

Yates, F. (1947) The analysis of dates from all possible reciprocal crosses between a set of parental lines. Heredity, 1, 287-301.

APPENDIX 1

```
 'REFERENCE' DIALLEL
'CAPTION'   ''
   P is the number of parental lines
   Sets up size parameters
   I is the index for male parent,
   J is the index for female parent,
   D is the index for copies of data,
     variates,factors,and calculations
     for PARents,Pure Versus Crosses,
     RECIPricals,DUMMY.
   Crosses (i,j) and (j,i) are at level
       ix(i-1)/2+j of RECIP.
''
 'SET' P=9     'SCALAR' N,N1,N2
 'CALC' N=2*P*P  : N1=P*P  : N2=P*(P-1)/2
 'RUN'
 'UNITS' UNSN=1...N
 'VARIATE' I=(P!(1...P))2 : J=((1...P)P)2 : D=N1!(1,2)
 'VARIATE' PARV,PVCV,RECIPV,DUMV
 'FACTOR' PARENTSSP : PVCS2  : RECIPSN2 : DUMMYSN1 : COPYS2
 'CALC' PARV=I*(D.EQ.1) + J*(D.EQ.2)
 'CALC' PVCV=(I.EQ.J) + 2*(I.NE.J)
 'CALC' RECIPV=(I.GE.J)*(I*(I-1)/2 + J) + (I.LT.J)*(J*(J-1)/2 + I)
 'CALC' DUMV=(I-1)*P + J
 'GROUP' PARENTS,PVC,RECIP,DUMMY,COPY=INTPT(PARV,PVCV,RECIPV,DUMV,D)
 'HEADING'  H = ''
  THIS GIVES FOUR TIMES THE NUMBERS IN JINKS AND MATHER (page 256) ''
 'DESCRIBE'     YT $ ; H
 'BLOCKS' RECIP/DUMMY
 'TREATMENTS' PARENTS*PVC + PARENTS*COPY
 'VARIATE' Y$N1  'VARIATE' YT
 'READ/PRIN=DEM' Y  'EQUATE' YT=Y
 'DESCIBE' YT$5
 'ANOVA/PR=00010' YT
 'RUN'
```

```
77.8 53.4 79.6 69.6 50.2 59.6 71.4 67.6 50.6
47.8 54.1 50.0 46.2 43.0 52.4 46.8 41.2 40.4
68.8 53.2 97.6 59.1 50.0 63.0 72.2 48.8 52.0
72.2 47.0 62.4 68.2 46.8 58.7 54.4 44.6 50.0
53.0 46.4 52.0 51.0 53.2 55.0 54.4 40.4 48.4
56.8 48.2 60.6 63.8 48.3 54.0 55.4 44.8 49.6
73.8 49.4 83.6 67.8 60.2 59.6 74.0 48.8 58.2
53.6 38.6 55.6 44.2 38.4 37.6 45.4 30.6 43.6
50.6 46.6 49.8 48.0 45.0 42.6 54.8 38.0 50.8
'EOD'
Y     MNMINMAX        54.2616        30.6000        97.6000     81  VALUES     0 MISSING

1
 ***** ANALYSIS OF VARIANCE *****

 VARIATE: YT
   THIS GIVES FOUR TIMES THE NUMBERS IN JINKS AND MATHER (page 256)

 SOURCE OF VARIATION                DF          SS       SS%          MS        VR

 RECIP STRATUM
   PARENTS                           8   1.582E  4     72.85    1.978E  3    23.238
   PVC                               1   1.294E  3      5.96    1.294E  3    15.203
   PARENTS.PVC                       8   1.146E  3      5.27    1.432E  2     1.683
   RESIDUAL                         27   2.298E  3     10.58    8.511E  1
 TOTAL                              44   2.056E  4     94.66    4.673E  2

 RECIP.DUMMY STRATUM
   PARENTS.COPY                      8   5.420E  2      2.50    6.775E  1     3.068
   RESIDUAL                         28   6.183E  2      2.85    2.208E  1
 TOTAL                              36   1.160E  3      5.34    3.223E  1

 RECIP.DUMMY.*UNITS* STRATUM
   PARENTS                           8   0.000E  0      0.00    0.000E  0
   COPY                              1   0.000E  0      0.00    0.000E  0
   PARENTS.COPY                      8   0.000E  0      0.00    0.000E  0
   RESIDUAL                         64   0.000E  0      0.00    0.000E  0
 TOTAL                              81   0.000E  0      0.00    0.000E  0

 GRAND TOTAL                       161   2.172E  4    100.00

 GRAND MEAN                      54.26172
 TOTAL NUMBER OF OBSERVATIONS         162
 ***** INFORMATION SUMMARY *****

 MODEL TERM                          EF  NON-ORTHOGONAL TERMS

 RECIP STRATUM
   PARENTS                       0.500
   PARENTS.PVC                   0.875  PARENTS

 RECIP.DUMMY STRATUM
   PARENTS.COPY                  0.500

 RECIP.DUMMY.*UNITS* STRATUM
   PARENTS                       0.500  RECIP
   PARENTS.COPY                  0.500  RECIP.DUMMY

 'CLOSE'
```

APPENDIX 2

```
 'REFERENCE'          ROW_COLUMN_ASSOCIATION
 ''
  This file is RCA
 ''
 'CAPTION'            ''
  The row and column associations are now calculated by grouping the levels
  of factor T ''
 'UNITS'              $ 60
 'INTEGER'            INT1 = -11,-1,2,-5,3,6,-8,4,7,9,-10
     :                INT2 = 1,2,3,-4,5,6,-7,8,-9,-10,-11
```

```
'FACTOR'        R $ 5
    :           C $ 5
    :           T $ 10 = 3(1),2(2),2(3),2,2(4),3,4,7,(6,5)2,
                5,8,9,8,2(7),6,9,2(10),8,10,9
    :           DUMMY $ 30 = (1...30)2
    :           BLOCK $ 6 = (1...6)10
'RUN'

'GROUP'         R = GROUP(T ; INT1)
    :           C = GROUP(T ; INT2)
'RUN'


'FACTOR'        TREAT $ 10 = T,T
    :           PF1 $ 5 = R,C
'VARIATE'       Y = (3(-3),2(-2,-1),-2,2(0),-1,0,1,0,-1,0,2(-1),
                1,2,3(1),0,2,2(3),1,3,2)2
    :           BLCKT = (-5,-3,-1,1,3,5)10
    :           INTERACT = (5(0),2(2),0,2(-2),(2,-2)2,0,-2,
                5(0),2(2),-2,6(0))2
'PRINT/P'       DUMMY,BLOCK,TREAT,PF1,Y,BLCKT,INTERACT $ 10
'RUN'
```

DUMMY	BLOCK	TREAT	PF1	Y	BLCKT	INTERACT
1	1	1	2	-3	-5	0
2	2	1	2	-3	-3	0
3	3	1	2	-3	-1	0
4	4	2	3	-2	1	0
5	5	2	3	-2	3	0
6	6	3	4	-1	5	2
7	1	3	4	-1	-5	2
8	2	2	3	-2	-3	0
9	3	4	5	0	-1	-2
10	4	4	5	0	1	-2
11	5	3	4	-1	3	2
12	6	4	5	0	5	-2
13	1	7	5	1	-5	2
14	2	6	4	0	-3	-2
15	3	5	3	-1	-1	0
16	4	6	4	0	1	-2
17	5	5	3	-1	3	0
18	6	5	3	-1	5	0
19	1	8	4	1	-5	0
20	2	9	5	2	-3	0
21	3	8	4	1	-1	0
22	4	7	5	1	1	2
23	5	7	5	1	3	2
24	6	6	4	0	5	-2
25	1	9	5	2	-5	0
26	2	10	5	3	-3	0
27	3	10	5	3	-1	0
28	4	8	4	1	1	0
29	5	10	5	3	3	0
30	6	9	5	2	5	0
1	1	1	1	-3	-5	0
2	2	1	1	-3	-3	0
3	3	1	1	-3	-1	0
4	4	2	1	-2	1	0
5	5	2	1	-2	3	0
6	6	3	1	-1	5	2
7	1	3	1	-1	-5	2
8	2	2	1	-2	-3	0
9	3	4	1	0	-1	-2
10	4	4	1	0	1	-2

11	5	3	1	-1	3	2
12	6	4	1	0	5	-2
13	1	7	2	1	-5	2
14	2	6	2	0	-3	-2
15	3	5	2	-1	-1	0
16	4	6	2	0	1	-2
17	5	5	2	-1	3	0
18	6	5	2	-1	5	0
19	1	8	3	1	-5	0
20	2	9	3	2	-3	0
21	3	8	3	1	-1	0
22	4	7	2	1	1	2
23	5	7	2	1	3	2
24	6	6	2	0	5	-2
25	1	9	3	2	-5	0
26	2	10	4	3	-3	0
27	3	10	4	3	-1	0
28	4	8	3	1	1	0
29	5	10	4	3	3	0
30	6	9	3	2	5	0

```
'BLOCKS'            BLOCK/DUMMY
'TREATMENT'         TREAT//PF1
'CALCULATE'         Y = Y + BLCKT
   :                Y = Y + INTERACT
'ANOVA/PR=03013'    Y
'RUN'

1
 ***** ANALYSIS OF VARIANCE *****

 VARIATE: Y

 SOURCE OF VARIATION                DF          SS        SS%          MS        VR

 BLOCK STRATUM
   TREAT                             5   6.712E  2      72.33    1.342E  2
 TOTAL                               5   6.712E  2      72.33    1.342E  2

 BLOCK.DUMMY STRATUM
   TREAT                             9   2.568E  2      27.67    2.853E  1
   RESIDUAL                         15   0.000E  0       0.00    0.000E  0
 TOTAL                              24   2.568E  2      27.67    1.070E  1

 BLOCK.DUMMY.*UNITS* STRATUM
   TREAT                             4   0.000E  0       0.00    0.000E  0
   RESIDUAL                         26   0.000E  0       0.00    0.000E  0
 TOTAL                              30   0.000E  0       0.00    0.000E  0

 GRAND TOTAL                        59   9.280E  2     100.00

 GRAND MEAN                          0.00
 TOTAL NUMBER OF OBSERVATIONS          60

 ***** INFORMATION SUMMARY *****

 MODEL TERM                          EF  NON-ORTHOGONAL TERMS

 BLOCK STRATUM
   TREAT                          0.200

 BLOCK.DUMMY STRATUM
   PF1                            0.375
   TREAT                          0.800  BLOCK

 BLOCK.DUMMY.*UNITS* STRATUM
   PF1                            0.625  BLOCK.DUMMY

 ***** TABLES OF EFFECTS *****

 VARIATE: Y
```

```
*** BLOCK STRATUM ***

 TREAT                    EFFECTS:              REP         6  ESE          *

        TREAT        1        2        3        4        5        6        7        8        9       10
                -15.00     1.67     7.00     6.33    11.67     3.00     0.33    -8.33    -5.00    -1.67
*** BLOCK.DUMMY STRATUM ***

 PF1                      EFFECTS:              REP        12  ESE      0.000

        PF1          1        2        3        4        5
                 -4.00    -2.00     0.00     2.00     4.00

 TREAT                    EFFECTS:              REP         6  ESE      0.000

       TREAT         1        2        3        4        5        6        7        8        9       10
                  0.00     0.00     2.00    -2.00    -0.00    -2.00     2.00    -0.00    -0.00    -0.00

*** BLOCK.DUMMY.*UNITS* STRATUM ***

 PF1                      EFFECTS:              REP        12  ESE      0.000

        PF1          1        2        3        4        5
                  0.00     0.00     0.00     0.00     0.00

***** TABLES OF MEANS *****

VARIATE: Y

    GRAND MEAN     0.00

       TREAT         1        2        3        4        5        6        7        8        9       10

         PF1         2        2        3        3        3        3        4        4        4        5
                  0.00     0.00     2.00    -2.00    -0.00    -2.00     2.00    -0.00    -0.00    -0.00

***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE                 TREAT
---------------------------
REP                       6
CED                   0.000

***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM                  DF          SE         CV%

BLOCK                     0           *          *
BLOCK.DUMMY              15       0.000          *
BLOCK.DUMMY.*UNITS*      26       0.000          *


    'CLOSE'
```

APPENDIX 3

```
'REFERENCE/NID=100,NUNN=100'        TRIPLE_RL
'CAPTION'                 ''
This program uses an example of a triple rectangular lattice taken from
 'The Design and Analysis of Experiments' by Kempthorne (page 522).
''
'LINES'                   10
'UNITS'                   $ 36
'FACTOR'                  R $ 3 = 12(1...3)
    :                     T4 $ 3
    :                     B,T1,T2,T3 $ 4
    :                     T $ 12
'READ/PRIN=DE,FORM=P'     Y,T,T1,T2,T3,B,T4
'RUN'
 8.9  2 1 3 2 1 2
10.0  3 1 4 3 1 3
11.6  1 1 2 4 1 1
 9.4  7 3 1 4 3 2
 9.3  9 3 4 2 3 1
10.2  8 3 2 1 3 3
 9.6  4 2 1 3 2 1
11.4  5 2 3 4 2 3
10.0  6 2 4 1 2 2
11.8 10 4 1 2 4 3
11.7 12 4 3 1 4 1
13.1 11 4 2 3 4 2
11.6  2 1 3 2 3 2
12.4 12 4 3 1 3 1
10.0  5 2 3 4 3 3
 9.6  7 3 1 4 1 2
10.0 10 4 1 2 1 3
 8.4  4 2 1 3 1 1
10.5  1 1 2 4 2 1
11.0  8 3 2 1 2 3
10.0 11 4 2 3 2 2
11.5  3 1 4 3 4 3
12.6  9 3 4 2 4 1
11.8  6 2 4 1 4 2
11.2  2 1 3 2 2 2
12.1  9 3 4 2 2 1
 9.7 10 4 1 2 2 3
 7.8  3 1 4 3 3 3
 9.6 11 4 2 3 3 2
10.6  4 2 1 3 3 1
 8.7 12 4 3 1 1 1
 9.1  8 3 2 1 1 3
 8.0  6 2 4 1 1 2
 8.7  1 1 2 4 4 1
 7.5  7 3 1 4 4 2
 8.3  5 2 3 4 4 3
'EOD'
'PRINT/P'                 Y,R,B,T,T1,T2,T3,T4 $ 10.2,7(10)
'LINES'                   10
'BLOCKS'                  R/B
'TREATMENTS'              T//(T1+T2+T3+T4)
'CAPTION'                 ''
This analysis uses the data and gives an approximate table of means ''
'ANOVA'                   Y
'RUN'
```

Y	R	B	T	T1	T2	T3	T4
8.90	1	1	2	1	3	2	2
10.00	1	1	3	1	4	3	3
11.60	1	1	1	1	2	4	1
9.40	1	3	7	3	1	4	2
9.30	1	3	9	3	4	2	1
10.20	1	3	8	3	2	1	3
9.60	1	2	4	2	1	3	1
11.40	1	2	5	2	3	4	3
10.00	1	2	6	2	4	1	2
11.80	1	4	10	4	1	2	3
11.70	1	4	12	4	3	1	1
13.10	1	4	11	4	2	3	2
11.60	2	3	2	1	3	2	2
12.40	2	3	12	4	3	1	1
10.00	2	3	5	2	3	4	3
9.60	2	1	7	3	1	4	2
10.00	2	1	10	4	1	2	3
8.40	2	1	4	2	1	3	1
10.50	2	2	1	1	2	4	1
11.00	2	2	8	3	2	1	3
10.00	2	2	11	4	2	3	2
11.50	2	4	3	1	4	3	3
12.60	2	4	9	3	4	2	1
11.80	2	4	6	2	4	1	2
11.20	3	2	2	1	3	2	2
12.10	3	2	9	3	4	2	1
9.70	3	2	10	4	1	2	3
7.80	3	3	3	1	4	3	3
9.60	3	3	11	4	2	3	2
10.60	3	3	4	2	1	3	1
8.70	3	1	12	4	3	1	1
9.10	3	1	8	3	2	1	3
8.00	3	1	6	2	4	1	2
8.70	3	4	1	1	2	4	1
7.50	3	4	7	3	1	4	2
8.30	3	4	5	2	3	4	3

'his analysis uses the data and gives an approximate table of means

**** ANALYSIS OF VARIANCE *****

'ARIATE: Y

OURCE OF VARIATION	DF	SS	SS%	MS	VR
STRATUM	2	16.1072	21.75	8.0536	8.329
.B STRATUM					
T	9	36.9025	49.84	4.1003	
OTAL	9	36.9025	49.84	4.1003	4.240
.B.*UNITS* STRATUM					
T	11	8.4630	11.43	0.7694	0.796
RESIDUAL	13	12.5703	16.98	0.9669	
OTAL	24	21.0333	28.41	0.8764	
RAND TOTAL	35	74.0430	100.00		

RAND MEAN	10.21
OTAL NUMBER OF OBSERVATIONS	36

**** INFORMATION SUMMARY *****

)DEL TERM	EF	NON-ORTHOGONAL TERMS
.B STRATUM		
T1	0.407	
T2	0.323	T1
T3	0.099	T1 T2

```
R.B.*UNITS* STRATUM
  T1                          0.593  R.B
  T2                          0.556  R.B  T1
  T3                          0.494  R.B  T1  T2

ALIASED MODEL TERMS
T


***** TABLES OF MEANS *****

VARIATE: Y


    GRAND MEAN     10.21

*** FOLLOWING TABLE OF MEANS IS ONLY APPROXIMATE ***

  T        1       2       3       4       5       6       7       8       9      10      11

 T1        1       1       1       2       2       2       3       3       3       4       4
 T2        2       3       4       1       3       4       1       2       4       1       2
 T3        4       2       3       3       4       1       4       1       2       2       3
 T4        1       2       3       1       3       2       2       3       1       3       2
       11.24   10.03    9.25    9.47    9.98    9.42   10.18   11.29   10.69    9.75   10.60

  T       12

 T1        4
 T2        3
 T3        1
 T4        1
       10.66

 ***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

 TABLE                        T
 ---------------------------
 REP                          3
 SED                      1.160
 EXCEPT WHEN COMPARING MEANS WITH SAME LEVEL(S) OF:
   T1                     0.991
   T2                     0.979
   T3                     0.954
   T4                     1.088


 ***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

 STRATUM                     DF          SE          CV%

 R                            2       0.819          8.0
 R.B                          0           *            *
 R.B.*UNITS*                 13       0.983          9.6


'UNITS'          UNIT $ 108 = 1...108
'VARIATE'        NY = (Y)3
'FACTOR'         NTP $ 4 = T1,T2,T3
    :            COPY  $ 3 = 36(1...3)
    :            NT4 $ 3 = (T4)3
    :            DUMMY $ 36 = (1...36)3
    :            NT $ 12 = (T)3
    :            NR $ 3 = (R)3
    :            NB $ 4 = (B)3
'DEVALUE'        Y,T1,T2,T3,T,R,B
'DESCRIBE'       NY $ 3
'LINES'          10
'TREATMENT'      NT//(NTP*COPY+NT4)
'BLOCKS'         NR/NB/DUMMY
'CAPTION'        ''
This analysis uses copies of the data and in the analysis of variance table
gives three times the numbers in the above analysis ''
'ANOVA/PR=03013'  NY
'RUN'
```

This analysis uses copies of the data and in the analysis of variance table gives three times the numbers in the above analysis

1

***** ANALYSIS OF VARIANCE *****

VARIATE: NY

SOURCE OF VARIATION	DF	SS	SS%	MS	VR
NR STRATUM	2	4.832E 1	21.75	2.416E 1	
NR.NB STRATUM					
NT	9	1.107E 2	49.84	1.230E 1	
TOTAL	9	1.107E 2	49.84	1.230E 1	
NR.NB.DUMMY STRATUM					
NT	11	2.539E 1	11.43	2.308E 0	0.796
RESIDUAL	13	3.771E 1	16.98	2.901E 0	
TOTAL	24	6.310E 1	28.41	2.629E 0	
NR.NB.DUMMY.*UNITS* STRATUM					
NT	11	0.000E 0	0.00	0.000E 0	
RESIDUAL	61	0.000E 0	0.00	0.000E 0	
TOTAL	72	0.000E 0	0.00	0.000E 0	
GRAND TOTAL	107	2.221E 2	100.00		

GRAND MEAN 10.214
TOTAL NUMBER OF OBSERVATIONS 108

***** INFORMATION SUMMARY *****

MODEL TERM	EF	NON-ORTHOGONAL TERMS	
NR.NB STRATUM			
NTP	0.012		
NTP.COPY	0.198		
NR.NB.DUMMY STRATUM			
NTP	0.099	NR.NB	
NTP.COPY	0.247	NR.NB	
NR.NB.DUMMY.*UNITS* STRATUM			
NTP	0.889	NR.NB	NR.NB.DUMMY
NTP.COPY	0.556	NR.NB	NR.NB.DUMMY

ALIASED MODEL TERMS
NT

***** TABLES OF EFFECTS *****

VARIATE: NY

*** NR.NB STRATUM ***

NTP EFFECTS: REP 27 ESE

NTP	1	2	3	4
	-7.625	3.575	-1.025	5.075

NTP.COPY EFFECTS: REP 9 ESE

COPY / NTP	1	2	3
1	0.969	-1.356	0.388
2	-1.456	-1.531	2.988
3	-1.881	1.494	0.388
4	2.369	1.394	-3.762

** NR.NB.DUMMY STRATUM ***

NTP EFFECTS: REP 27 ESE 1.0430

NTP	1	2	3	4
	-0.725	0.800	0.062	-0.137

NTP.COPY EFFECTS: REP 9 ESE 1.1425

COPY / NTP	1	2	3
1	-0.160	-0.640	0.800
2	-1.055	1.570	-0.515
3	1.025	-0.145	-0.880
4	0.190	-0.785	0.595

* NR.NB.DUMMY.*UNITS* STRATUM ***

NTP EFFECTS: REP 27 ESE 0.0000

NTP	1	2	3	4
	0.000	0.000	0.000	0.000

COPY EFFECTS: REP 36 ESE 0.0000

COPY	1	2	3
	0.000	0.000	0.000

TP.COPY EFFECTS: REP 9 ESE 0.0000

COPY / NTP	1	2	3
1	0.000	0.000	0.000
2	0.000	0.000	0.000
3	0.000	0.000	0.000
4	0.000	0.000	0.000

***** TABLES OF MEANS *****

VARIATE: NY

GRAND MEAN 10.214

NT	1	2	3	4	5	6	7	8	9	10	11
NTP	2	2	3	2	3	2	3	2	3	2	3
COPY	2	2	2	2	2	2	2	2	2	2	2
NT4	1	2	3	1	3	2	2	3	1	3	2
	10.517	10.058	10.067	10.517	10.067	10.058	10.058	10.067	10.517	10.067	10.058

NT	12
NTP	3
COPY	2
NT4	1
	10.517

***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE	NT
REP	9
SED	0.4014

***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM	DF	SE	CV%
NR	2	0.8192	8.0
NR.NB	0	*	*
NR.NB.DUMMY	13	0.9833	9.6
NR.NB.DUMMY.*UNITS*	61	0.0000	0.0

'CLOSE'

R Thompson
Agricultural Research Council
Unit of Statistics
University of Edinburgh
James Clerk Maxwell Building
The King's Buildings
Mayfield Road
Edinburgh EH9 3JZ
U.K.

J. Statist. Comput. Simul., 1986, Vol. 24, pp. 215–230
0094-9655/86/2404-0215 $18.50/0
© 1986 Gordon and Breach, Science Publishers, Inc.
Printed in Great Britain

Estimation of Variance Components: What is Missing in the *EM* Algorithm?

R. THOMPSON and K. MEYER

Animal Breeding Research Organisation and Institute of Animal Genetics, Kings Buildings, Edinburgh, Scotland

(*Received June 4, 1985*)

The *EM* algorithm is a frequently advocated algorithm for the estimation of variance components. A faster converging algorithm is developed using alternative parameterizations based on the analysis of variance. The procedure is exemplified using designs with two and three variance components and with multivariate designs using parameter values relevant to animal breeding data.

KEY WORDS: Variance components, *EM* algorithm, *REML*, analysis of variance.

1. THE MODEL

We consider estimation in linear models specified by

$$\mathbf{y}=\mathbf{X}\boldsymbol{\alpha}+\sum_{i=1}^{c}\mathbf{Z}_i\mathbf{b}_i+\mathbf{e} \tag{1.1}$$

where $\mathbf{y}$ is a $n\times 1$ vector of observed responses, $\mathbf{X}$ and $\mathbf{Z}_i$ $(i=1,\ldots,c)$ are known matrices of size $n\times p$ and $n\times q_i$, α is a $p\times 1$ vector of fixed effects, and each $\mathbf{b}_i$ is a $q_i\times 1$ vector of random effects distributed independently as $N(0,\sigma_i^2\mathbf{I})$ and $\mathbf{e}$ is a $n\times 1$ vector of error

terms distributed as $N(0, \sigma^2\mathbf{I})$. Suppose elements of $\mathbf{Z}_i$ have values zero or one and if each row of $\mathbf{Z}_i$ has only one non-zero element. Eq. (1.1) can then be thought of as an ANOVA mixed model.

We consider in this paper computation of variance components for this model by restricted maximum likelihood (*REML*) a procedure based on maximizing the likelihood (L) of error contrasts (Patterson and Thompson, 1971). If the data structures are unbalanced, iterative schemes are needed. We have found that iterative schemes using the expected values of the second differentials of L (Harville (1977), Patterson and Thompson (1971)) are often computationally feasible and can converge fairly quickly. In some cases, results on partitioned matrices (Thompson 1977a), Robinson, Thompson and Digby (1982)) can give a more efficient computing strategy. But there are data structures, for example Meyer (1983), where the calculation of second differentials is computationally expensive.

An alternative approach, for example Dempster, Selwyn, Patel and Roth (1984) and Laird (1982) is to use the *EM* algorithm. This is a general purpose algorithm, discussed by Dempster, Rubin and Laird (1977), that can be used to compute maximum likelihood estimates for a large number of statistical models, that can be thought of as using the first differentials of L. Thompson (1977b) used asymptotic formulae of Dempster *et al.* (1977) to show theoretically that the algorithm can be slow to converge for a one-way classification, especially for values of variance components common in animal breeding studies. This slow convergence has been found in other designs, for instance Thompson (1979).

However for most balanced designs there is a unique partition of the sum of squares into orthogonal parts due to the various factors in the model. It is then a simple matter to transform the mean squares for random factors, which estimate stratum variances (Nelder, 1965) to give estimates of the variance components. The fact that these explicit estimates exist suggests that iterative schemes based on first differentials might be available that give estimates for balanced designs in one iteration.

Animal breeding data that we deal with is seldom exactly balanced because of biological variation, for instance in litter size and sex of offspring, but experimenters usually have a target balanced design in mind. Under these circumstances one would hope a scheme which is optimal for balanced designs, i.e. yield estimates in one round of

iteration, would also be more efficient for unbalanced designs, i.e. converge faster than existing schemes. This idea is investigated for several designs and parameter values appropriate for animal breeding work.

2. MODELS WITH TWO VARIANCE COMPONENTS

In this section we consider model (1.1) with $p=1$ and $c=1$, including a between group component σ_1^2 and a within component σ^2 for several designs of increasing complexity.

a) Balanced one-way classification

The model can be written as:

$$y_{ij}=\mu+b_{1i}+e_{ij} \quad (i=1,\ldots,s, j=1,\ldots,m) \tag{2.1}$$

or in the formulation of (1.1) $p=1$ and $\mathbf{X}$ is a $sm\times 1$ matrix of ones, $\mu=\alpha$, $q_1=s$ and $sm=n$ and $\mathbf{b}_1$ and $\mathbf{e}$ have elements b_i and e_{ij}. The log-likelihood of error contrasts (L) is given by

$$2L=-s(m-1)\ln\sigma^2-(s-1)\ln(\sigma^2+m\sigma_1^2)-W/\sigma^2-B/(\sigma^2+m\sigma_1^2) \tag{2.2}$$

where W and B are the within and between group sum of squares. The EM algorithm is based on treating μ, b_i and e_{ij} as "missing data" and arguing that if e_{ij} and b_i were known then one can form sufficient statistics

$$t=\Sigma e_{ij}^2 \quad \text{and} \quad t_1=\Sigma b_i^2 \tag{2.3}$$

and estimate σ^2 and σ_1^2 from

$$sm\sigma^2=t \quad \text{and} \quad s\sigma_1^2=t_1. \tag{2.4}$$

Because t and t_1 are not known Dempster *et al.* (1977) suggest that t and t_1 in (2.3) be replaced by their posterior expectation given σ_1^2 and σ^2 (an E or expectation step). As (2.4) gives a maximum

likelihood estimator this can be thought of as an M (or maximization step). Hence the two steps combine to give an EM algorithm. This gives rise to an iterative scheme increasing L in each iteration. An alternative, more direct, derivation is to differentiate L with respect to the "natural" parameters $-1/2\sigma^2$ and $-1/2\sigma_1^2$ and set the derivatives to zero (Dempster *et al.*, 1984). For model (2.1) it can be shown that one step of the EM algorithm changes σ^2 and σ_1^2 to $\hat{\sigma}^2$ and $\hat{\sigma}_1^2$ where

$$sm\hat{\sigma}^2 = W + \sigma^4 B^* + s\sigma^2 \tag{2.5}$$

and

$$s\hat{\sigma}_1^2 = s\sigma_1^2 + \sigma_1^4 m B^* \tag{2.6}$$

where

$$B^* = [B/(\sigma^2 + m\sigma_1^2) - (s-1)]/(\sigma^2 + m\sigma_1^2).$$

Table I gives an analysis of variance for such data. This gives estimates of σ^2 and σ_1^2 from

$$s(m-1)\sigma^2 = W \quad \text{and} \quad (s-1)(\sigma^2 + m\sigma_1^2) = B \tag{2.7}$$

which is of the same form as (2.4) with degrees of freedom, $s(m-1)$ and $(s-1)$, replacing the number of effects, sm and s. Further $\sigma^2 + m\sigma_1^2$ in (2.7) replaces σ_1^2 in (2.4). This suggests using σ^2, the residual variance, and $\theta_{1m} = \sigma_1^2 + \sigma^2/m$, the variance of a group mean, as an alternative parameterization, or more generally σ^2 and $\theta_{1k} = \sigma_1^2 + \sigma^2/k$, noting that as $k \to \infty$, $\theta_{1k} \to \sigma_1^2$.

TABLE I

Analysis of variance.

Source of variation	d.f.	Sum of squares	Expected mean square
Between groups	$s-1$	B	$\sigma^2 + m\sigma_1^2$
Within groups	$s(m-1)$	W	σ^2

Differentiating (2.1) with respect to $-1/2\sigma^2$ and $-1/2\theta_{1k}$ gives estimating equations:

$$s(m-1)\sigma^2 = W + (1-m/k)\sigma^4 B^* \tag{2.8}$$

$$(s-1)\theta_{1k} = (s-1)\theta_{ik} + \theta_{1k}^2 mB^*. \tag{2.9}$$

These equations reduce to (2.7) when $k=m$ and are a minor rearrangement of (2.5) and (2.6) as $k \to \infty$.

For a design with $m=25$ and $s=20$, Figures 1 and 2 show estimates of σ_1^2 found using (2.8) and (2.9) for $k=\infty$ and $k=25$, respectively. B and W were derived using $\sigma_1^2=0.125$ and $\sigma^2=0.875$ in Figure 1, and $\sigma_1^2=0.0125$ and $\sigma^2=0.9875$ in Figure 2. Whilst values of $\sigma_1^2/(\sigma^2+\sigma_1^2)$ are small, these are in the range of values often found in animal breeding studies. In both cases starting values of $\sigma_1^2=0.0625$ and $\sigma^2=0.9375$ were used. These figures emphasize the slow rate of convergence of the *EM* algorithm using the original parameters σ^2 and σ_1^2 ($k=\infty$) especially when σ_1^2/σ^2 is small.

b) Unbalanced classification

In order to investigate the effect of lack of balance, suppose we have p supergroups each with a different mean α_i and that within the ith supergroup we have a balanced structure of s_i groups, each of m_i observations ($i=1,\ldots,p$). Then if B_{1i} is the sum of squares between groups within supergroup i and W is the pooled within group sum of squares, estimates can be derived analogous to (2.8) and (2.9) based on σ^2 and θ_{1k} from

$$[\Sigma s_i(m_i-1)]\hat{\sigma}^2 = W + \sigma^4[\Sigma[(1-m_i/k)B_i^*] \tag{2.10}$$

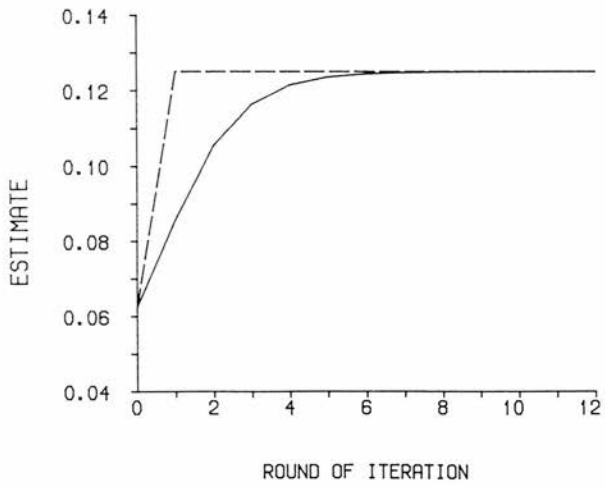
and

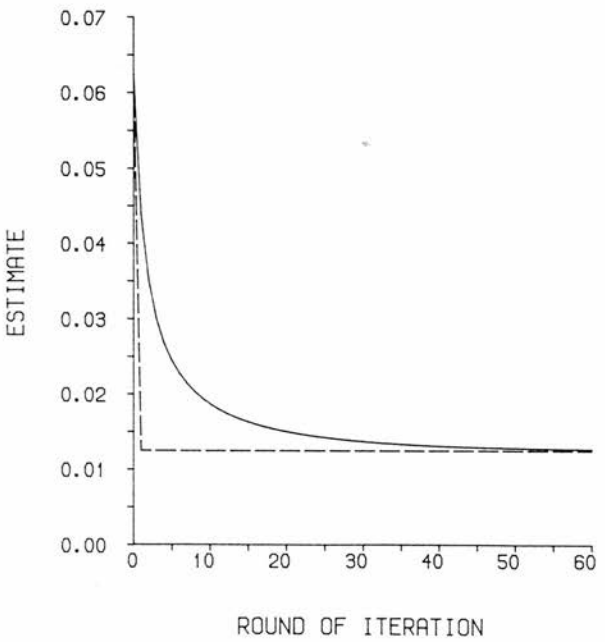
$$[\Sigma(s_i-1)]\hat{\theta}_{1k} = [\Sigma(s_i-1)]\theta_{ik} + \theta_{ik}^2[\Sigma m_i B_i^*] \tag{2.11}$$

where

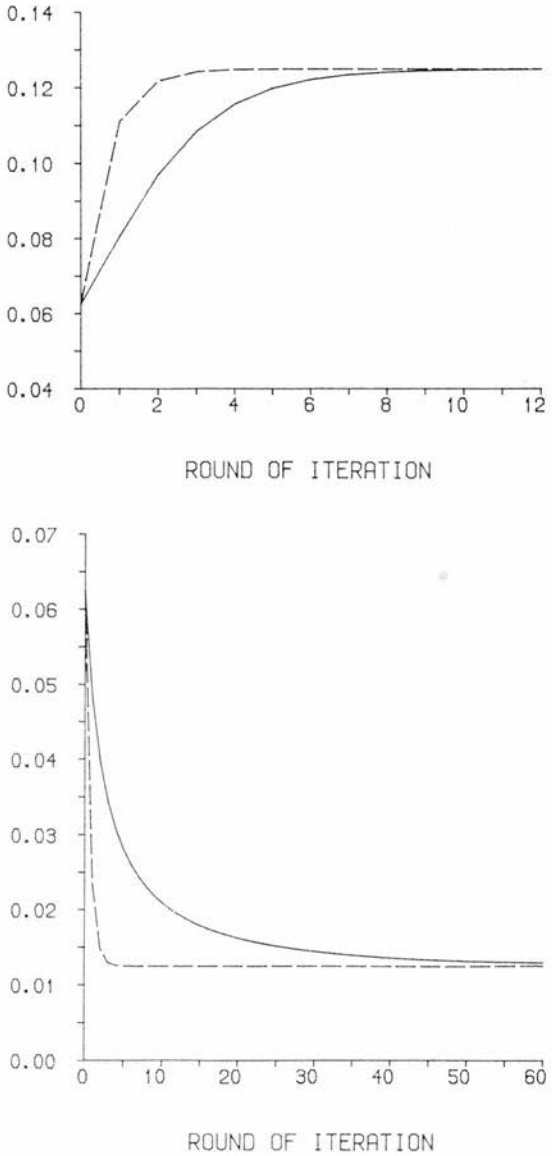
$$B_i^* = [B_i/(\sigma^2+m_i\sigma_1^2) - (s_i-1)]/(\sigma^2+m_i\sigma_1^2).$$

Figures 3 and 4 show estimates of σ_1^2 over rounds of iteration obtained using (2.10) and (2.11). The design examined had $p=5$





Figures 1 and 2 Values of σ_1^2 from an iterative scheme for a balanced design $k=\infty$ =solid curve; $k=25$ =dotted curve.



FIGURES 3 and 4 Values of σ_1^2 from an iterative scheme for an unbalanced design, $k = \infty$ = solid curve, $k = 25$ = dotted curve.

supergroups with four groups in each supergroup. The respective group sizes were $m_i = 15$, 20, 25, 30 and 35 with a mean of 25. Starting values and values of σ^2 and σ_1^2 used to derive B_i and W were as for Figures 1 and 2. Again the untransformed EM algorithm ($k = \infty$) is contrasted to the reparameterization with $k = 25$, the mean family size. For the latter estimates of σ_1^2 converge quickly, whilst the rate of convergence for $k = \infty$ is similar to that for the balanced design (see Figures 1 and 2).

c) Mixed model

Slightly more complicated formulae are needed if the model of analysis (1.1) includes estimable fixed effects $p(>1)$.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}_1\mathbf{b}_1 + \mathbf{e}.$$

Then weighted least squares estimates of α can be found from

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_1 \\ \mathbf{Z}_1'\mathbf{X} & \mathbf{Z}_1'\mathbf{Z}_1 + \mathbf{I}\gamma \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}_1'\mathbf{y} \end{bmatrix}.$$

With $\gamma = \sigma^2/\sigma_1^2$, or rearranging (2.11)

$$\begin{bmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xz} \\ \mathbf{C}_{zx} & \mathbf{C}_{zz} \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}_1'\mathbf{y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}.$$

It can be shown by manipulating formulae given by Harville (1977) and Patterson and Thompson (1971), that an iterative scheme analogous to (2.5) and (2.6) is given by

$$(n-p)\sigma^2 = \mathbf{e}'\mathbf{e} - \boldsymbol{\beta}_1'\boldsymbol{\beta}_1\sigma^4/k\sigma_1^4 + q\sigma^2 + q\sigma^4/k\sigma_1^2 - tr(\mathbf{C}_{zz})\theta_{1k}\sigma^4/\sigma_1^4, \tag{2.12}$$

$$q\hat{\theta}_{1k} = q(\theta_{ik} - \theta_{1k}^2/\sigma_1^2) + (\theta_{1k}/\sigma_1^2)^2[\boldsymbol{\beta}_1'\boldsymbol{\beta}_1 + \sigma^2 tr(\mathbf{C}_{zz})], \tag{2.13}$$

where $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{Z}\boldsymbol{\beta}$, and q, as defined in (1.1), is the number of levels of the random effect.

This method was applied to a reproductive study recently considered by Dempster *et al.* (1984). Estimates of σ_1^2 and σ^2 for different choices of k are given in Table II showing the improvement in reducing k from infinity.

TABLE II

Iterative values of parameters for reproductive data.

	Estimates of σ_1^2			Estimates of σ^2		
Iteration	$k=\infty$	$k=11.93$	$k=9$	$k=\infty$	$k=11.93$	$k=9$
0	0.064289	0.064289	0.064289	0.251014	0.257014	0.251014
1	0.076860	0.093968	0.100165	0.162718	0.161212	0.160725
2	0.084268	0.093814	0.097581	0.163220	0.162825	0.162847
3	0.094367	0.097305	0.097419	0.162922	0.162801	0.162802
4	0.096299	0.097384	0.097402	0.162843	0.162802	
5	0.097004	0.097397	0.097400	0.162816		
6	0.097258	0.097399		0.162807		
7	0.097344	0.097400		0.152803		
8	0.097381			0.162802		
9	0.097393					
10	0.097397					

Patterson and Thompson (1971) and Dempster *et al.* (1984) note that the matrix inversion needed to find $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_1$ in (2.11) for each round of (2.12) and (2.13) can be avoided if a spectral decomposition of $\mathbf{Z}_1'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Z}=\mathbf{PDP}'$ is used. The matrix $\mathbf{P}$ is an orthogonal matrix $\mathbf{D}$ is a diagonal matrix with elements λ_i of which r are non-zero elements. This can be used to construct an analysis of variance with r sum of squares b_i each with one degree of freedom, formed from the squares of the vector $\mathbf{P}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}$ with expectation $\sigma^2+\lambda_i\sigma_1^2$. There is also a within group of squares $W=\mathbf{e}'\mathbf{e}$ found using $\gamma=0$ to evaluate $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_1$. This analysis of variance is similar to that for the unbalanced classification (Section 2b) and so (2.9) and (2.10) could be used replacing the residual degrees of freedom $\Sigma s_i(m_i-1)$ in (2.9) by n-rank $[\mathbf{XZ}_1]$ and replacing (s_i-1) by 1 and m_i by λ_i.

3. HIERARCHICAL DESIGN WITH THREE VARIANCE COMPONENTS

A common animal breeding design is an hierarchical design when each father is mated to several mothers and several offspring are raised and measured from each mating. There is interest in estimating the father (σ_2^2), mother (σ_1^2) and residual (σ^2) variance components, i.e. $c=2$ in model (1.1). Suppose there are p supergroups each with a balanced structure. In the ith supergroup observations have mean subscript i and there are measurements on $s_i d_i m_i$ animals. Each of s_i fathers is mated to d_i mothers and m_i offspring from each mating are measured. If the design was balanced with $s_i=s$, $d_i=d$ and $m_i=m$ for $i=1,\ldots,p$, a hierarchical analysis of variance could be constructed and the mean squares between fathers (F), between mothers within fathers (M) and within mothers (W) could be used to estimate $\sigma^2+d\sigma_1^2+md\sigma_2^2$, $\sigma^2+m\sigma_1^2$ and σ^2, respectively. The argument of Section 2 suggests for this unbalanced design using $\theta_{1k_1}=\sigma_1^2+\sigma^2/k_1$ and $\theta_{2k_2}=\sigma_2^2+(\theta_{1k_1})/k_2=\sigma_2^2+\sigma_1^2/k_2+\sigma^2/k_1k_2$ as parameters.

An iterative solution schemes based on σ^2, θ_{1k_1} and θ_{2k_2} is given by

$$[s_i d_i(m_i-1)]\hat{\sigma}^2 = W+\sigma^4\Sigma(1-m_i/k_1)[M_i^*+F_i^*] \tag{3.1}$$

$$[\Sigma s_i(d_i-1)]\hat{\theta}_{1k_1}=[\Sigma s_i(d_i-1)]\theta_{1k_1}+\theta_{1k_1}^2\Sigma[m_iM_i^*+(1-d_i/k_2)m_iF_i^*] \tag{3.2}$$

$$[\Sigma(s_i-1)]\hat{\theta}_{2k_2}=[\Sigma(s_i-1)]\theta_{2k_2}+\theta_{2k_2}^2\Sigma[m_id_iF_i^*] \tag{3.3}$$

where

$$F_i^*=(F_i/(\sigma^2+m_i\sigma_1^2+m_id_i\sigma_2^2)-(s_i-1))/(\sigma^2+m_i\sigma_1^2+m_id_i\sigma_2^2)$$

$$M_i^*=(M_i/(\sigma^2+m_i\sigma_1^2)-s_i(d_i-1))/(\sigma^2+m_i\sigma_1^2).$$

F_i denotes the sums of squares between fathers, and M_i the sum of squares between mothers within fathers in the ith group, and W is the within mother sum of squares. For a balanced design $m_i=m$, $d_i=d$ and $s_i=s$. Then (3.1), (3.2) and (3.3) with $k_1=m$ and $k_2=d$ reduce to $psd(m-1)\sigma^2=W$, $ps(d-1)m\theta_{1m}=\Sigma M_i$, $p(s-1)m\theta_{2d}=\Sigma F_i$, i.e. as for the balanced one-way classification estimates converge in one round of iteration.

A design was constructed with 12 supergroups and $s_i = 5$ for each supergroup. To generate m_i and d_i, four values of d_i, 2, 4, 6 and 8, and three values of m_i, 3, 5 and 7, were chosen and each combination of d_i and m_i was used in one supergroup. Figures 5 and 6 show estimates for σ_1^2 (the slowest converging parameter) using $k_1 = k_2 = \infty$ and $k_1 = k_2 = 5$ in (3.1), (3.2) and (3.3).

W, F_i and M_i were derived using $\sigma^2 = 0.75$ $\sigma_1^2 = \sigma_2^2 = 0.125$ (Figure 5) and $\sigma^2 = 0.975$ $\sigma_1^2 = \sigma_2^2 = 0.0125$ (Figure 6) and starting values were $\sigma^2 = 0.875$ and $\sigma_1^2 = \sigma_2^2 = 0.0625$. Again the figures show the advantage of reducing k_1 and k_2 from infinity.

4. MULTIVARIATE VARIANCE COMPONENTS

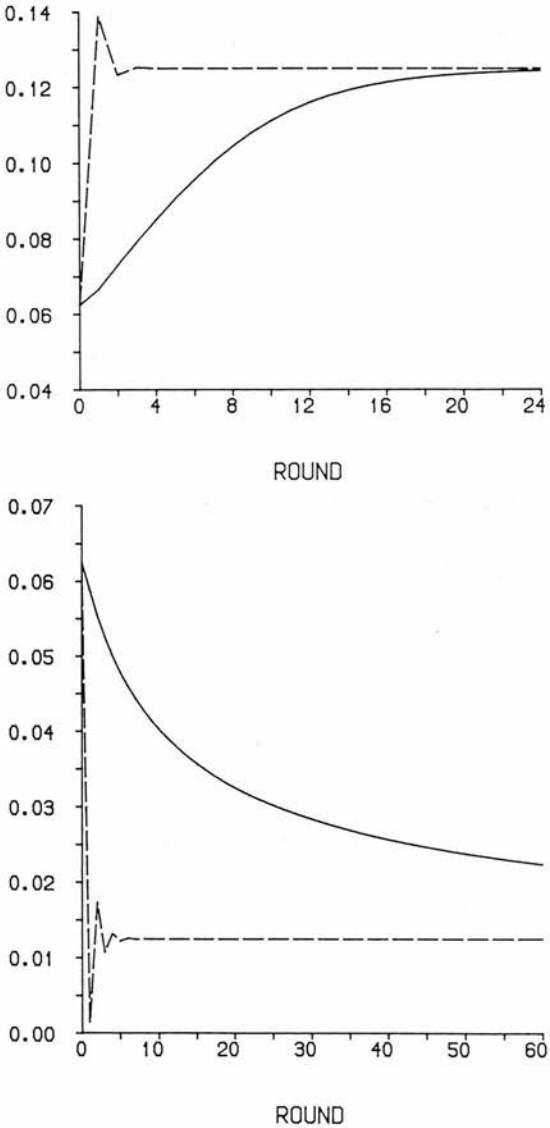
In this section we consider a multivariate generalization motivated by a problem considered by Meyer (1983). We suppose t variates can be measured and that a model of the form (1.1) is appropriate for each variate with $c = 1$. We assume that the random effects for the ith and jth traits have covariance $\sigma_{1ij} I$ and the errors terms have covariance $\sigma_{ij}\mathbf{I}$. There are then $t(t+1)$ variance parameters σ_{1ij} and σ_{ij} $(j \leqq 1,\ i = 1, \ldots, t)$ which can be written as two $t \times t$ symmetric matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_1$. One simple design, a generalization of Section 2b, is to assume that observations are in p supergroups and that in the ith supergroup t_i variates are measured on m_i individuals in s_i groups. The mean effect for the jth variate in the ith group is α_{ij}. For example with $p = 2$ supergroups and $t = 2$ variates measured, we might have the first variate measured on m individuals in each of s_i groups in the first supergroup and both variates measured on m_2 individuals in each of s_2 groups in the second supergroup.

The log-likelihood of error contrasts for the ith supergroup (L_i) is given by

$$2L_i = -(s_i - 1)\ln\left|\boldsymbol{\Sigma}_i + m_i\boldsymbol{\Sigma}_{1i}\right| - s_i(m_i - 1)\ln\left|\boldsymbol{\Sigma}_i\right| - \operatorname{trace}(\boldsymbol{\Sigma}_i^{-1}\mathbf{W}_i)$$

$$- \operatorname{trace}\left((\boldsymbol{\Sigma}_i + m_i\boldsymbol{\Sigma}_{1i})^{-1}\mathbf{B}_i\right)$$

where $\mathbf{W}_i$ and $\mathbf{B}_i$ are the within and between sum of squares and cross-products for the t_i variates measured in ith group, and $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Sigma}_{1i}$ are the subset of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_1$ relevant to the measurements in the



FIGURES 5 and 6 Values of σ_1^2 from an iterative scheme for an unbalanced hierarchical design, $k_1 = k_2 \infty =$ solid curve; $k_1 = k_2 = 5$ dotted curve.

ith supergroup. The argument in the previous section suggests using $\boldsymbol{\Sigma}$ and $\boldsymbol{\theta}_{1k}$ instead of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_1$ as parameters where $\boldsymbol{\theta}_{1k} = \mathbf{K} \odot \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}$ and the matrix $\mathbf{K}$ is a symmetrical matrix with elements K_{ij} $(j \leqq i,\ i = 1, \ldots, t)$ and $\odot$ is an operator that multiplies two matrices term by term, i.e. the ijth element of $\mathbf{A} \odot \mathbf{B}$ has (i, j) element $A_{ij}B_{ij}$.

An iterative scheme for $\boldsymbol{\Sigma}$ and $\boldsymbol{\theta}_{1k}$ is then

$$\mathbf{D} \odot \boldsymbol{\Sigma} = D \odot \boldsymbol{\Sigma} + \boldsymbol{\Sigma}[dL/d\boldsymbol{\Sigma}]\boldsymbol{\Sigma} \tag{4.1}$$

$$\mathbf{D}_1 \odot \boldsymbol{\theta}_{1k} = \mathbf{D}_1 \odot \boldsymbol{\theta}_{1k} + \boldsymbol{\theta}_{1k}[dL/\boldsymbol{\theta}_{1k}]\boldsymbol{\theta}_{1k} \tag{4.2}$$

where $dL/d\boldsymbol{\phi}$ $(\boldsymbol{\phi} = \boldsymbol{\Sigma}, \boldsymbol{\theta}_{1k})$ is a symmetric matrix defined by diagonal elements $dL/d\phi_{ii}$ and off diagonal elements $(\frac{1}{2})\ dL/d\phi_{ij}$ $(i < j)$. Correspondingly, $\mathbf{D}$ and $\mathbf{D}_1$ are symmetric matrices with (i, j)th element equal to the degrees of freedom within and between groups, for the ith and jth trait respectively.

For the example above with two supergroups then

$$\mathbf{D} = \begin{bmatrix} s_1(m_1 - 1) + s_2(m_2 - 1) & s_2(m_2 - 1) \\ s_2(m_2 - 1) & s_2(m_2 - 1) \end{bmatrix} \quad \text{and}$$

$$\mathbf{D}_1 = \begin{bmatrix} s_1 - 1 + s_2 - 1 & s_2 - 1 \\ s_2 - 1 & s_2 - 1 \end{bmatrix}.$$

As a multivariate example, we take a design with

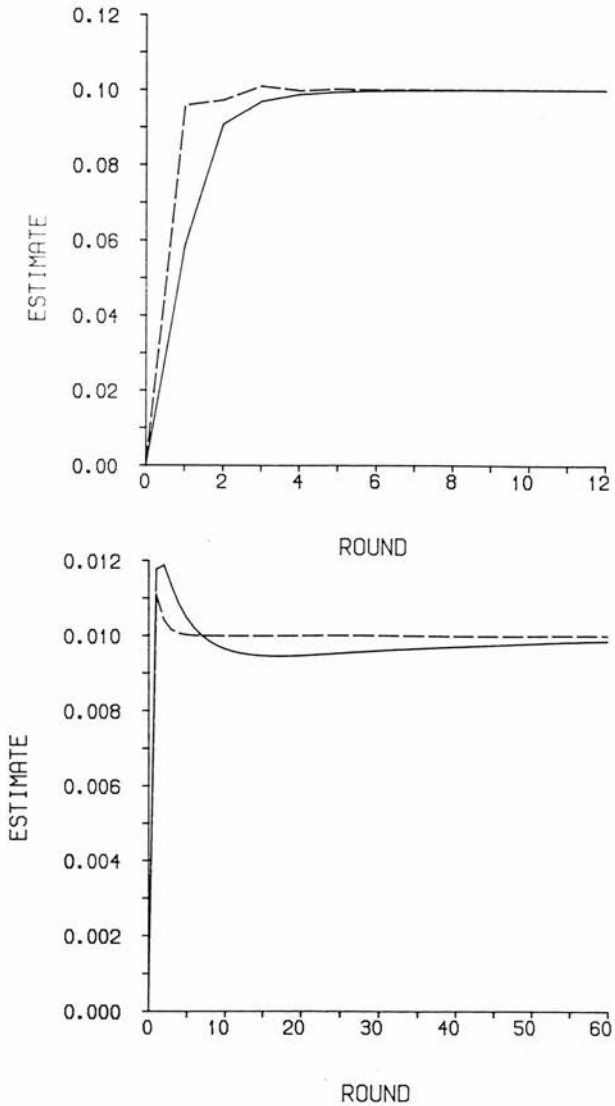
$$s_1 = 50,\ s_2 = 50,\ m_1 = 10,\ m_2 = 40$$

and calculate W_i and B_i using

i) $$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.125 & 0.100 \\ 0.100 & 0.125 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.875 & 0.400 \\ 0.400 & 0.875 \end{bmatrix}$$

and

ii) $$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.0125 & 0.0100 \\ 0.0100 & 0.0125 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.9875 & 0.4900 \\ 0.4900 & 0.9875 \end{bmatrix}$$



FIGURES 7 and 8 Values of σ_{112} from an iterative scheme for an unbalanced multivariate design.

$$k=\begin{bmatrix}\infty & \infty\\ \infty & \infty\end{bmatrix}=\text{solid curve;} \qquad k=\begin{bmatrix}50 & 40\\ 40 & 40\end{bmatrix}=\text{dotted curve.}$$

and starting values

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.0625 & 0.0000 \\ 0.0000 & 0.0625 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.9375 & 0.0000 \\ 0.0000 & 0.9375 \end{bmatrix}.$$

These estimates and parameter values are a multivariate extension of those used in Section 2. Figures 7 and 8 show graphs of estimates of σ_{12}, using (4.1) and (4.2) for cases (i) and (ii), using values for

$$\mathbf{K} = \begin{bmatrix} \infty & \infty \\ \infty & \infty \end{bmatrix} \quad \text{and} \quad \mathbf{K} = \begin{bmatrix} 50 & 40 \\ 40 & 40 \end{bmatrix}.$$

5. CONCLUSION

It has been shown that in a variety of settings, consideration of balanced designs suggest a reparameterization that improves convergence. These schemes have been constructed directly from the first differentials of L but, no doubt, could have been derived by considering linear functions, of $\boldsymbol{\alpha}$, $\mathbf{b}_1$ and $\mathbf{e}$, such as $\mathbf{b}_1 - (1/k_1)\mathbf{Z}_1'\mathbf{Z}_1\mathbf{e}$ and treat them as "missing values" in the *EM* algorithm. There are other variance component problems with non-normal data (Laird (1978), Harville and Mee (1984)) where this approach is probably useful.

One property of the *EM* algorithm is that the variance component estimates will always be greater than or equal to zero (Harville (1977)). By contrast the scheme outlined can lead to negative estimates. However the *EM* algorithm becomes slower to converge as variance components tend to zero and we have found it more practical to use our scheme and then constrain or attenuate the parameter estimates if we wish to impose restrictions on the estimates.

References

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood with incomplete data via the *EM* algorithm. *Journal of the Royal Statistical Society Series B*, **39**, 1–38.

Dempster, A. P., Selwyn, M. R., Patel, C. M. and Roth, A. J. (1984). Statistical and computational aspects of mixed model analysis. *Applied Statistics*, **33**, 203–214.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72,** 320–340.

Harville, D. A. and Mee, R. W. (1984). A mixed-model procedure for analysing ordered categorial data. *Biometrics*, **40,** 393–408.

Laird, N. M. (1978). Empirical Bayes methods for two-way contingency tables. *Biometrika*, **65,** 581–590.

Laird, N. M. (1982). Computation of variance components using the *EM* algorithm. *J. Statist. Comput. Simul.*, **14,** 295–303.

Meyer, K. (1983). Maximum likelihood procedures for estimating genetic parameters for later lactations of dairy cattle. *J. Dairy Science*, **66,** 1988–1997.

Nelder, J. A. (1965). The analysis of randomized experiments with orthogonal block structure: Block structure and the null analysis of variance. *Proc. Roy. Soc. A.*, **283,** 147–162.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58,** 545–554.

Robinson, D. L., Digby, P. G. N. and Thompson, R. (1982). *REML* a program for the analysis of non-orthogonal data by restricted maximum likelihood. *COMPSTAT 82, Part II*, Physica-Verlag, Vienna.

Thompson, R. (1977a). Discussion of the paper by Dempster *et al. Journal of Royal Statistical Society*, **39,** 34.

Thompson, R. (1977b). The estimation of heritability with unbalanced data. *Biometrics*, **33,** 485–495.

Thompson, R. (1979). Sire evaluation. *Biometrics*, **35,** 339–354.

11

A program for the analysis of non-orthogonal data by restricted maximum likelihood

by

D.L. Robinson, R. Thompson and P.G.N. Digby

REML - A Program for the Analysis of Non-orthogonal data by Restricted Maximum Likelihood.

D L Robinson, R Thompson and P G N Digby.

In any unbalanced experiment with more than one factor, good estimates of treatments must take into account the lack of balance in the levels of the other factors. With fixed effects, this may be achieved by regression. If there are also unbalanced blocking factors which are not orthogonal to treatment factors, for example in incomplete block designs, then efficient treatment estimates can only be found when the variances of these blocking factors are known, enabling recovery of inter-block information.

An appropriate analytical technique is that of Restricted Maximum Likelihood (Patterson and Thompson, 1971) which gives unbiased estimates identical to those from an analysis of variance in the case of balanced data. As well as the efficient estimation of fixed effects. The magnitude of sources of variation can be identified and effects related to the random factors can be predicted.

These are important; for example in sire evaluation it is usual to assume that a daughter's yield had a fixed herd-year-season effect and a random sire effect. From this inherently unbalanced dataset prediction of the random sire effect is required, so that chosen sires can later be used extensively, on the basis of their daughter's yields. The magnitude and sources of variation are also important so that testing strategies can be evaluated, and optimal schemes used.

Such an analysis has been programmed in standard Fortran, based on a modification of the algorithm proposed by Thompson (1977). The usual regression equations are reformulated as $(X'X+G^{-1})\underline{a}=X'\underline{y}$, where X is obtained from the design matrix for fixed and random effects, and $\underline{y}$ is a vector of observed values. Random effects are incorporated by the diagonal matrix G^{-1} , whose ith entry is 0 if the ith parameter a_i in $\underline{a}$ is fixed, and $1/v$ if a_i is a random effect from a population with variance v . An iterative technique, similar to Fisher's scoring technique, is used to solve the likelihood equations. Initial values of variance components are formed, or supplied by the user, and used to estimate the parameter vector $\underline{x}$; new variance components are then found from the derivatives of the likelihood equations.

Since X'X can be written in a form that is largely block diagonal, most of the parameter estimates can be formed sequentially using small submatrices which occupy the same storage locations, only the relevant parameters being retained.

The estimates for the likelihood equations are accumulated at the same time. Computing time is saved since only small matrices need to be inverted, and storage requirements are lessened considerably.

A wide range of linear models may be fitted, with no restriction on the number of fixed or random factors or covariates. Input is by simple directives specifying the factors, variates and covariates in the model to be fitted. Storage is allocated dynamically, so the directives may appear in almost any order. The data can be read sequentially or in parallel, or any combination of the two, and in free or fixed format.

The user can specify the amount of output required, from simple, easy to read, tables of selected estimates of effects, optionally with average standard errors, to a complete list of all the parameter estimates and tables of pairwise standard errors at every iteration. The former will be useful if there are factors with a large number of levels which are not of interest and would clutter the output, the latter may aid interpretation of the data. Often the most useful combination is to print (i) the components of variance evey iteration or two to check on convergence, and (ii) the required estimates of effects and variances after the final iteration. Approximate stratum variances based on a canonical decomposition of the information matrix are also available. In addition a full program monitor can be used to check on some or all of the calculations performed.

Jennrich and Sampson (1978) ask "What if anything can be done about the fact that general purpose maximum likelihood programs use a lot of computer storage placing rather severe size limitations on the size of problems that can be handled?". By the sequential formation of small submatrices occupying the same storage locations REML overcomes this problem and has fitted models to datasets with over 9,000 observations, some factors having ~~nearly~~ 1,400 levels.

References.

Jennrich, R I, and Sampson, P F (1978). Some problems faced in making a variance component algorithm into a general mixed model program. Proceedings of the 11th annual symposium on the interface between computer science and statistics, Raleigh, 1978, 56-63.

Patterson, H D, and Thompson, R (1971). Recovery of inter-block information when block sizes are unequal. Biometrics 58, 445-554.

Thompson, R (1977). The estimation of heritability with unbalanced data. Biometrics 33, 485-495.

12

Diallel crosses, partially balanced inccomplete block designs with triangular association schemes and rectangular lattices

by

R. Thompson

DIALLEL CROSSES, PARTIALLY BALANCED INCOMPLETE BLOCK DESIGNS WITH TRIANGULAR ASSOCIATION SCHEMES AND RECTANGULAR LATTICES

INTRODUCTION

What have these designs in common? Apart from esoteric names, all three have a high degree of symmetry and yet cannot be completely analysed by the Genstat 'ANOVA' directive. The treatment structure for the first design cannot be specified, the 'ANOVA' algorithm finds the second design unbalanced and treatment means for the third are only 'approximate'. Of course, other methods of analysis are possible and macros are available to analyse (i) diallel crosses using regression procedures (N.G. Alvey and O. Mayo, personal communications) and (ii) rectangular lattices using 'ANOVA' on the dual of the design (E.R. Williams and D. Ratcliff, CSIRO Macro).

I suggest instead a technique which aids in the analysis and, I think, the understanding of all three designs (and no doubt others). It is based on the idea of repeating the data several times and generating treatment and block factors with each copy of the data. The treatment and block factors are chosen so that the repeated data has a 'balanced' structure and that the required analysis is a meaningful subset (usually the only one) of the analysis of the extended data. Thompson and Baker (1981) found a similar idea useful in embedding several types of data into a generalised linear model framework.

I will demonstrate the technique for each design in turn and then discuss the disadvantages of the method and point out how the Genstat 'ANOVA' algorithm might be improved.

DIALLEL CROSSES

In plant breeding work on p different parental lines, often, all reciprocal crosses can be made and the p^2 progeny used for the comparison of parental lines. Various models and analyses have been suggested (Yates (1947), Hayman (1954), Griffing (1956), Jinks and Mather (1971)). I first consider a simple model that exhibits the difficulty in fitting these classes of models. I take the case where the progeny mean, when a male of line i is crossed with a female of line j, is y_{ij}. One might think of fitting a factorial model

$$y_{ij} = \textit{mean + male line (i) effect + female line (j) effect}$$

and this can be easily done in 'ANOVA'. However it might be reasonable to assume the male and female effects of line i to be similar and so want to fit a model

$$y_{ij} = \textit{mean + line (i) effect + line (j) effect.} \qquad (1)$$

However, each observation has 2 contributions from the line factor and so it is difficult to specify the treatment structure. Suppose, however, that we have two observations yt_{ij1} and yt_{ij2} with expectations

$$yt_{ij1} = \textit{mean + line (i) effect}\ , \quad yt_{ij2} = \textit{mean + line (j) effect.}$$

Each such pair yt_{ij1} and yt_{ij2} can be put in a different level of a factor DUMMY (for example put yt_{ij1} and yt_{ij2} into level $(i+(j-1)xp)$. For $p=3$ the data, treatment and block factors will be

$$\begin{aligned}
Y &= (y_{11} \quad y_{12} \quad y_{13} \quad y_{21} \quad y_{22} \quad y_{23} \quad y_{31} \quad y_{32} \quad y_{33}) \\
YT &= (yt_{111} \quad yt_{121} \quad yt_{131} \quad yt_{211} \quad yt_{221} \quad yt_{231} \quad yt_{311} \quad yt_{321} \quad yt_{331} \\
&\quad\ yt_{112} \quad yt_{122} \quad yt_{132} \quad yt_{212} \quad yt_{222} \quad yt_{232} \quad yt_{312} \quad yt_{322} \quad yt_{332}) \\
&= (y_{11} \quad y_{12} \quad y_{13} \quad y_{21} \quad y_{22} \quad y_{23} \quad y_{31} \quad y_{32} \quad y_{33} \\
&\quad\ y_{11} \quad y_{12} \quad y_{13} \quad y_{21} \quad y_{22} \quad y_{23} \quad y_{31} \quad y_{32} \quad y_{33}) \\
LINE &= (1 \quad 1 \quad 1 \quad 2 \quad 2 \quad 2 \quad 3 \quad 3 \quad 3 \\
&\quad\ 1 \quad 2 \quad 3 \quad 1 \quad 2 \quad 3 \quad 1 \quad 2 \quad 3) \\
DUMMY &= (1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \\
&\quad\ 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9).
\end{aligned}$$

Then an analysis of the p^2 x 2 values yt_{ijk} with LINE as treatment factor and DUMMY as a blocking factor will use 1 in the DUMMY stratum and, hence, give the required analysis in the DUMMY stratum. As noted by Wilkinson (1971) the resulting line effects have efficiency $(p-2)/(p-1)$.

Other more complicated models might be fitted. In Appendix 1 the code is given for Hayman's (1954) analysis using data in Mather and Jinks (1971, p. 256). This analysis includes comparisons of mean parental performance with mean progeny performance (i.e. y_{ii} versus y_{ij} $(j=i)$ or 'pure' versus 'cross') and of line parental performance with line cross performance and fits a model similar to (1) to the reciprocal difference $(y_{ij}-y_{ji})$.

PARTIALLY BALANCED INCOMPLETE BLOCKS WITH TRIANGULAR ASSOCIATION SCHEMES (P.B.I.B.T.)

These designs (for example Clatworthy, 1973) have $t = n(n-1)/2$ treatments. The association scheme can be obtained from a symmetric n x n array A with the diagonal blank and each treatment occurring twice - once above and once below the diagonal. Treatments in the same row or column are first associates. For example, the design with $t=10$, $b=6$, $k=5$, $r=3$, with treatments

1, 3, 7, 8, 9	in the first block
1, 2, 6, 9, 10	in the second block
1, 4, 5, 8, 10	in the third block
2, 4, 6, 7, 8	in the fourth block
2, 3, 5, 7, 10	in the fifth block and
3, 4, 5, 6, 9	in the sixth block

is a P.B.I.B.T..

The design has array A

*	1	2	3	4
1	*	5	6	7
2	5	*	8	9
3	6	8	*	10
4	7	9	10	*

and this indicates that 2, 3, 4, 5, 6 and 7 are the first associates of 1 and 3, 6, 8, 4, 7 and 9 are first associates of 10. As A is symmetric, the information on the association scheme can be found from the triangular array above (or below) the diagonal and hence the designs are said to have a triangular association scheme. The designs cannot be analysed directly by 'ANOVA' but they can be made balanced by imposing a pseudo-treatment structure. Because of the association scheme it is natural to use the rows

and columns of A to generate some pseudo-treatment structure. Suppose that treatment i (in the lower triangle of A) is in row r_i and column c_i. Then T.P. Speed has shown that P.B.I.B.T. designs have general balance with respect to the pseudo-treatment scheme

$$\begin{aligned} i\text{-th treatment effect} &= r_i\text{-th pseudo-treatment effect} \\ &+ c_i\text{-th pseudo-treatment effect} \\ &+ r_i, c_i \text{ pseudo-treatment interaction effect} \end{aligned}$$

with $(n-1)$ degrees of freedom for the pseudo-treatment effects and $(n-2)(n-1)/2$ degrees of freedom for the interaction effects.

This model is similar to the diallel model and can be fitted in a similar way. We make 2 copies of the data and use the rows of A to give the levels of the pseudofactor for the first copy and the columns of A to give the levels of the pseudofactor for the second. For example, the treatments

1 2 3 4 5 6 7 8 9 10

can be replaced by

2 3 4 5 3 4 5 4 5 5

with the first copy of the data and by

1 1 1 1 2 2 2 3 3 4

with the second copy of the data. Appendix 2 gives the Genstat code for the analysis of variance for this design, with some artificial data.

One problem is that Genstat does not know that each treatment is associated with 2 levels of the pseudofactor and it is not clear how the levels are calculated when deriving tables of means. For instance with the example above, Genstat associates levels *2, 2, 3, 3, 3, 3, 4, 4, 4, 5* with the treatments *1-10*. The level for treatment i is the integer part of $[(r_i+c_i+1)/2]$. In most other circumstances, the association of two levels of a pseudofactor with one treatment combination will only occur by mistake and it is intended to trap this in future releases of Genstat! However, correct tables of means and standard errors can easily be constructed using 'EXTRACT'.

A macro is available from the author to generate the levels of the pseudo-factors from treatment and block-factors for P.B.I.B.T. designs when A is not known, for example, if the treatments are randomised and no note is kept of the original order.

RECTANGULAR LATTICES

These designs, introduced by Harshbarger (1950) for $p(p-q)$ treatments in blocks of $(p-q)$, are a development of square lattices (Yates, 1936). Macros exist to analyse rectangular lattices but it is interesting to show how the method of the previous sections can also be used with these designs.

Consider a design for *12 = 4* x *3* treatments in 3 replications and block size of 3. A rectangular lattice design with these dimensions can be constructed from the following Graeco-Latin square

TP 1 \ *TP 2*	*1*	*2*	*3*	*4*
1	δ *13* *A*	α *1* *D*	β *2* *B*	γ *3* *C*
2	α *4* *C*	δ *14* *B*	γ *5* *D*	β *6* *A*
3	β *7* *D*	γ *8* *A*	δ *15* *C*	α *9* *B*
4	γ *10* *B*	β *11* *C*	α *12* *A*	δ *16* *D*

where the levels of *TP3* are the Latin letters *A*, *B*, *C*, *D* and of *TP4* are the Greek Letters α§ β§ γ§ δ. We associate with each cell a treatment, the off-diagonal cells are numbered from *1* to *12*, the diagonal from *13* to *16*. Then, using rows, columns and treatments to generate blocks, a 3-replicate rectangular lattice design can be derived by ignoring the diagonal elements of the square. (It will become apparent later why we have bothered to label the diagonal treatments as *13* to *16*.)

This is

Rep I	*1*	*2*	*3*	*Rep II*	*4*	*7*	*10*	*Rep III*	*6*	*8*	*12*
	4	*5*	*6*		*1*	*8*	*11*		*2*	*9*	*10*
	7	*8*	*9*		*2*	*5*	*12*		*3*	*4*	*11*
	10	*11*	*12*		*3*	*6*	*9*		*1*	*5*	*7*

This suggests using *TP1 + TP2 + TP3 + TP4* as a pseudofactor structure for *T* and, although this allows Genstat to produce the correct analysis of variance, the treatment effects are only 'approximate'. This is because *TP1*, *TP2* and *TP3* are not mutually orthogonal.

T.P. Speed has suggested an alternative subdivision of the pseudofactor space which allows an orthogonal subdivision. In our case, rather than think of *TP1*, *TP2*, *TP3* as representing 3 sets of 4 effects, we think of them as a two-way table indexed by *COPY* with 3 levels and *TP* with 4 levels, i.e.

		TP1	*TP2*	*TP3*
COPY		*1*	*2*	*3*
	1	*1*	*1*	*A*
TP	*2*	*2*	*2*	*B*
	3	*3*	*3*	*C*
	4	*4*	*4*	*D*

Speed has shown that *TP* and *TP.COPY* generate orthogonal subspaces (with 3 degrees of freedom and 6 degrees of freedom in our case). Note that the combinations of the pseudofactors *TP1*, *TP2*, *TP3* at level i of *TP* relate to the i-th diagonal of the Graeco-Latin square, so that the average effects of *TP* would give estimates to compare treatments *13* to *16* if the pseudo-factorial model was appropriate.

Hence, each treatment is now associated with 3 levels of *TP* and *TP.COPY*. As before, we can specify this model using 3 copies of the data and using *TPi* to give the level of *TP* with the ith copy of the data. Appendix 3 gives the Genstat code for a rectangular lattice taken from Kempthorne (1952, p.522). Again, there is a problem linking up treatment estimates with the 3 levels of TP and COPY.

DIFFICULTIES

There are three disadvantages of this method (i) multiple copies of, for instance, the yield variate are needed, (ii) effects and sums of squares in the dummy stratum are calculated, (iii) the linkage between treatments and pseudofactors is not complete.

G.N. Wilkinson has suggested that a simple modification of the algorithm should avoid at least the first two of the difficulties. He would allow the treatment factor length to be a multiple, say c, of the length of the analysed variate. When treatment totals are calculated, the program would go through the data c times and when sweeps are made, c terms would be taken from each data value. A similar technique could be used for deriving treatment estimates from pseudofactors.

REFERENCES

Clatworthy, W.H. (1973) Tables of two-associate-class partially balanced designs (US Dept of Commerce).

Griffing, B. (1956) A generalised treatment of the use of diallel crosses in quantative inheritance. Heredity, 10, 31-50.

Harshbarger, B. (1950) Triple rectangular lattices. Biometrics, 5, 1-13.

Hayman, B.I. (1954) The analysis of variance of diallel tables. Biometrics, 10, 235-244.

Jinks, J.L. and Mather, K. (1971) Biometric genetics. (Chapman and Hall).

Kempthorne, O. (1952) The design and analysis of experiments. J. Wiley and Sons, New York.

Thompson, R. and Baker, R.J. (1981) Composite link functions in generalised linear models. J.R.S.S., 30, 123-131.

Wilkinson, G.N. (1970) A general recursive algorithm for analysis of variance. Biometrika, 57, 19-46.

Yates, F. (1936) A new method of arranging variety trials involving a large number of varieties. Journal of Agricultural Science, 26, 424-455.

Yates, F. (1947) The analysis of dates from all possible reciprocal crosses between a set of parental lines. Heredity, 1, 287-301.

APPENDIX 1

```
 'REFERENCE' DIALLEL
'CAPTION'   ''
   P is the number of parental lines
   Sets up size parameters
   I is the index for male parent,
   J is the index for female parent,
   D is the index for copies of data,
     variates,factors,and calculations
     for PARents,Pure Versus Crosses,
     RECIPricals,DUMMY.
   Crosses (i,j) and (j,i) are at level
        ix(i-1)/2+j of RECIP.
''
 'SET' P=9    'SCALAR' N,N1,N2
 'CALC' N=2*P*P  : N1=P*P  : N2=P*(P-1)/2
 'RUN'
 'UNITS' UN$N=1...N
 'VARIATE' I=(P!(1...P))2 : J=((1...P)P)2 : D=N1!(1,2)
 'VARIATE' PARV,PVCV,RECIPV,DUMV
 'FACTOR' PARENTS$P : PVC$2  : RECIP$N2 : DUMMY$N1 : COPY$2
 'CALC' PARV=I*(D.EQ.1) + J*(D.EQ.2)
 'CALC' PVCV=(I.EQ.J) + 2*(I.NE.J)
 'CALC' RECIPV=(I.GE.J)*(I*(I-1)/2 + J) + (I.LT.J)*(J*(J-1)/2 + I)
 'CALC' DUMV=(I-1)*P + J
 'GROUP' PARENTS,PVC,RECIP,DUMMY,COPY=INTPT(PARV,PVCV,RECIPV,DUMV,D)
 'HEADING'  H = ''
  THIS GIVES FOUR TIMES THE NUMBERS IN JINKS AND MATHER (page 256) ''
 'DESCRIBE'    YT $ ; H
 'BLOCKS' RECIP/DUMMY
 'TREATMENTS' PARENTS*PVC + PARENTS*COPY
 'VARIATE' Y$N1  'VARIATE' YT
 'READ/PRIN=DEM' Y  'EQUATE' YT=Y
 'DESCIBE' YT$5
 'ANOVA/PR=00010' YT
 'RUN'
```

```
77.8 53.4 79.6 69.6 50.2 59.6 71.4 67.6 50.6
47.8 54.1 50.0 46.2 43.0 52.4 46.8 41.2 40.4
68.8 53.2 97.6 59.1 50.0 63.0 72.2 48.8 52.0
72.2 47.0 62.4 68.2 46.8 58.7 54.4 44.6 50.0
53.0 46.4 52.0 51.0 53.2 55.0 54.4 40.4 48.4
56.8 48.2 60.6 63.8 48.3 54.0 55.4 44.8 49.6
73.8 49.4 83.6 67.8 60.2 59.6 74.0 48.8 58.2
53.6 38.6 55.6 44.2 38.4 37.6 45.4 30.6 43.6
50.6 46.6 49.8 48.0 45.0 42.6 54.8 38.0 50.8
'EOD'
Y     MNMINMAX        54.2616        30.6000        97.6000     81  VALUES     0 MISSING

1
***** ANALYSIS OF VARIANCE *****

VARIATE: YT
  THIS GIVES FOUR TIMES THE NUMBERS IN JINKS AND MATHER (page 256)

SOURCE OF VARIATION                DF          SS        SS%          MS        VR

RECIP STRATUM
  PARENTS                           8   1.582E  4      72.85    1.978E  3     23.238
  PVC                               1   1.294E  3       5.96    1.294E  3     15.203
  PARENTS.PVC                       8   1.146E  3       5.27    1.432E  2      1.683
  RESIDUAL                         27   2.298E  3      10.58    8.511E  1
TOTAL                              44   2.056E  4      94.66    4.673E  2

RECIP.DUMMY STRATUM
  PARENTS.COPY                      8   5.420E  2       2.50    6.775E  1      3.068
  RESIDUAL                         28   6.183E  2       2.85    2.208E  1
TOTAL                              36   1.160E  3       5.34    3.223E  1

RECIP.DUMMY.*UNITS* STRATUM
  PARENTS                           8   0.000E  0       0.00    0.000E  0
  COPY                              1   0.000E  0       0.00    0.000E  0
  PARENTS.COPY                      8   0.000E  0       0.00    0.000E  0
  RESIDUAL                         64   0.000E  0       0.00    0.000E  0
TOTAL                              81   0.000E  0       0.00    0.000E  0

GRAND TOTAL                       161   2.172E  4     100.00

GRAND MEAN                      54.26172
TOTAL NUMBER OF OBSERVATIONS         162
***** INFORMATION SUMMARY *****

MODEL TERM                          EF  NON-ORTHOGONAL TERMS

RECIP STRATUM
  PARENTS                       0.500
  PARENTS.PVC                   0.875  PARENTS

RECIP.DUMMY STRATUM
  PARENTS.COPY                  0.500

RECIP.DUMMY.*UNITS* STRATUM
  PARENTS                       0.500  RECIP
  PARENTS.COPY                  0.500  RECIP.DUMMY

'CLOSE'
```

APPENDIX 2

```
'REFERENCE'          ROW_COLUMN_ASSOCIATION
''
 This file is RCA
''
'CAPTION'            ''
 The row and column associations are now calculated by grouping the levels
 of factor T ''
'UNITS'              $ 60
'INTEGER'            INT1 = -11,-1,2,-5,3,6,-8,4,7,9,-10
    :                INT2 = 1,2,3,-4,5,6,-7,8,-9,-10,-11
```

```
'FACTOR'        R $ 5
    :           C $ 5
    :           T $ 10 = 3(1),2(2),2(3),2,2(4),3,4,7,(6,5)2,
                5,8,9,8,2(7),6,9,2(10),8,10,9
    :           DUMMY $ 30 = (1...30)2
    :           BLOCK $ 6 = (1...6)10
'RUN'

'GROUP'         R = GROUP(T ; INT1)
    :           C = GROUP(T ; INT2)
'RUN'


'FACTOR'        TREAT $ 10 = T,T
    :           PF1 $ 5 = R,C
'VARIATE'       Y = (3(-3),2(-2,-1),-2,2(0),-1,0,1,0,-1,0,2(-1),
                1,2,3(1),0,2,2(3),1,3,2)2
    :           BLCKT = (-5,-3,-1,1,3,5)10
    :           INTERACT = (5(0),2(2),0,2(-2),(2,-2)2,0,-2,
                5(0),2(2),-2,6(0))2
'PRINT/P'       DUMMY,BLOCK,TREAT,PF1,Y,BLCKT,INTERACT $ 10
'RUN'
```

DUMMY	BLOCK	TREAT	PF1	Y	BLCKT	INTERACT
1	1	1	2	-3	-5	0
2	2	1	2	-3	-3	0
3	3	1	2	-3	-1	0
4	4	2	3	-2	1	0
5	5	2	3	-2	3	0
6	6	3	4	-1	5	2
7	1	3	4	-1	-5	2
8	2	2	3	-2	-3	0
9	3	4	5	0	-1	-2
10	4	4	5	0	1	-2
11	5	3	4	-1	3	2
12	6	4	5	0	5	-2
13	1	7	5	1	-5	2
14	2	6	4	0	-3	-2
15	3	5	3	-1	-1	0
16	4	6	4	0	1	-2
17	5	5	3	-1	3	0
18	6	5	3	-1	5	0
19	1	8	4	1	-5	0
20	2	9	5	2	-3	0
21	3	8	4	1	-1	0
22	4	7	5	1	1	2
23	5	7	5	1	3	2
24	6	6	4	0	5	-2
25	1	9	5	2	-5	0
26	2	10	5	3	-3	0
27	3	10	5	3	-1	0
28	4	8	4	1	1	0
29	5	10	5	3	3	0
30	6	9	5	2	5	0
1	1	1	1	-3	-5	0
2	2	1	1	-3	-3	0
3	3	1	1	-3	-1	0
4	4	2	1	-2	1	0
5	5	2	1	-2	3	0
6	6	3	1	-1	5	2
7	1	3	1	-1	-5	2
8	2	2	1	-2	-3	0
9	3	4	1	0	-1	-2
10	4	4	1	0	1	-2

11	5	3	1	-1	3	2
12	6	4	1	0	5	-2
13	1	7	2	1	-5	2
14	2	6	2	0	-3	-2
15	3	5	2	-1	-1	0
16	4	6	2	0	1	-2
17	5	5	2	-1	3	0
18	6	5	2	-1	5	0
19	1	8	3	1	-5	0
20	2	9	3	2	-3	0
21	3	8	3	1	-1	0
22	4	7	2	1	1	2
23	5	7	2	1	3	2
24	6	6	2	0	5	-2
25	1	9	3	2	-5	0
26	2	10	4	3	-3	0
27	3	10	4	3	-1	0
28	4	8	3	1	1	0
29	5	10	4	3	3	0
30	6	9	3	2	5	0

```
'BLOCKS'            BLOCK/DUMMY
'TREATMENT'         TREAT//PF1
'CALCULATE'         Y = Y + BLCKT
   :                Y = Y + INTERACT
'ANOVA/PR=03013'    Y
'RUN'

1
 ***** ANALYSIS OF VARIANCE *****

 VARIATE: Y

 SOURCE OF VARIATION                DF          SS        SS%          MS        VR

 BLOCK STRATUM
   TREAT                             5   6.712E  2     72.33   1.342E  2
 TOTAL                               5   6.712E  2     72.33   1.342E  2

 BLOCK.DUMMY STRATUM
   TREAT                             9   2.568E  2     27.67   2.853E  1
   RESIDUAL                         15   0.000E  0      0.00   0.000E  0
 TOTAL                              24   2.568E  2     27.67   1.070E  1

 BLOCK.DUMMY.*UNITS* STRATUM
   TREAT                             4   0.000E  0      0.00   0.000E  0
   RESIDUAL                         26   0.000E  0      0.00   0.000E  0
 TOTAL                              30   0.000E  0      0.00   0.000E  0

 GRAND TOTAL                        59   9.280E  2    100.00

 GRAND MEAN                          0.00
 TOTAL NUMBER OF OBSERVATIONS          60


 ***** INFORMATION SUMMARY *****

 MODEL TERM                          EF  NON-ORTHOGONAL TERMS

 BLOCK STRATUM
   TREAT                          0.200

 BLOCK.DUMMY STRATUM
   PF1                            0.375
   TREAT                          0.800  BLOCK

 BLOCK.DUMMY.*UNITS* STRATUM
   PF1                            0.625  BLOCK.DUMMY


 ***** TABLES OF EFFECTS *****

 VARIATE: Y
```

```
*** BLOCK STRATUM ***

 TREAT                    EFFECTS:              REP        6  ESE          *

      TREAT        1        2        3        4        5        6        7        8        9       10
             -15.00     1.67     7.00     6.33    11.67     3.00     0.33    -8.33    -5.00    -1.67
*** BLOCK.DUMMY STRATUM ***

 PF1                      EFFECTS:              REP       12  ESE      0.000

      PF1          1        2        3        4        5
             -4.00    -2.00     0.00     2.00     4.00


 TREAT                    EFFECTS:              REP        6  ESE      0.000

      TREAT        1        2        3        4        5        6        7        8        9       10
              0.00     0.00     2.00    -2.00    -0.00    -2.00     2.00    -0.00    -0.00    -0.00


*** BLOCK.DUMMY.*UNITS* STRATUM ***

 PF1                      EFFECTS:              REP       12  ESE      0.000

      PF1          1        2        3        4        5
              0.00     0.00     0.00     0.00     0.00


***** TABLES OF MEANS *****

VARIATE: Y


   GRAND MEAN     0.00

      TREAT        1        2        3        4        5        6        7        8        9       10

        PF1        2        2        3        3        3        3        4        4        4        5
              0.00     0.00     2.00    -2.00    -0.00    -2.00     2.00    -0.00    -0.00    -0.00


***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE                 TREAT
--------------------------
REP                       6
SED                   0.000


***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM                  DF          SE         CV%

BLOCK                     0           *          *
BLOCK.DUMMY              15       0.000          *
BLOCK.DUMMY.*UNITS*      26       0.000          *


     'CLOSE'
```

APPENDIX 3

```
'REFERENCE/NID=100,NUNN=100'        TRIPLE_RL
'CAPTION'           ''
This program uses an example of a triple rectangular lattice taken from
 'The Design and Analysis of Experiments' by Kempthorne (page 522).
''
'LINES'             10
'UNITS'             $ 36
'FACTOR'            R $ 3 = 12(1...3)
    :               T4 $ 3
    :               B,T1,T2,T3 $ 4
    :               T $ 12
'READ/PRIN=DE,FORM=P'   Y,T,T1,T2,T3,B,T4
'RUN'
 8.9  2 1 3 2 1 2
10.0  3 1 4 3 1 3
11.6  1 1 2 4 1 1
 9.4  7 3 1 4 3 2
 9.3  9 3 4 2 3 1
10.2  8 3 2 1 3 3
 9.6  4 2 1 3 2 1
11.4  5 2 3 4 2 3
10.0  6 2 4 1 2 2
11.8 10 4 1 2 4 3
11.7 12 4 3 1 4 1
13.1 11 4 2 3 4 2
11.6  2 1 3 2 3 2
12.4 12 4 3 1 3 1
10.0  5 2 3 4 3 3
 9.6  7 3 1 4 1 2
10.0 10 4 1 2 1 3
 8.4  4 2 1 3 1 1
10.5  1 1 2 4 2 1
11.0  8 3 2 1 2 3
10.0 11 4 2 3 2 2
11.5  3 1 4 3 4 3
12.6  9 3 4 2 4 1
11.8  6 2 4 1 4 2
11.2  2 1 3 2 2 2
12.1  9 3 4 2 2 1
 9.7 10 4 1 2 2 3
 7.8  3 1 4 3 3 3
 9.6 11 4 2 3 3 2
10.6  4 2 1 3 3 1
 8.7 12 4 3 1 1 1
 9.1  8 3 2 1 1 3
 8.0  6 2 4 1 1 2
 8.7  1 1 2 4 4 1
 7.5  7 3 1 4 4 2
 8.3  5 2 3 4 4 3
'EOD'
'PRINT/P'           Y,R,B,T,T1,T2,T3,T4 $ 10.2,7(10)
'LINES'             10
'BLOCKS'            R/B
'TREATMENTS'        T//(T1+T2+T3+T4)
'CAPTION'           ''
This analysis uses the data and gives an approximate table of means ''
'ANOVA'             Y
'RUN'
```

Y	R	B	T	T1	T2	T3	T4
8.90	1	1	2	1	3	2	2
10.00	1	1	3	1	4	3	3
11.60	1	1	1	1	2	4	1
9.40	1	3	7	3	1	4	2
9.30	1	3	9	3	4	2	1
10.20	1	3	8	3	2	1	3
9.60	1	2	4	2	1	3	1
11.40	1	2	5	2	3	4	3
10.00	1	2	6	2	4	1	2
11.80	1	4	10	4	1	2	3
11.70	1	4	12	4	3	1	1
13.10	1	4	11	4	2	3	2
11.60	2	3	2	1	3	2	2
12.40	2	3	12	4	3	1	1
10.00	2	3	5	2	3	4	3
9.60	2	1	7	3	1	4	2
10.00	2	1	10	4	1	2	3
8.40	2	1	4	2	1	3	1
10.50	2	2	1	1	2	4	1
11.00	2	2	8	3	2	1	3
10.00	2	2	11	4	2	3	2
11.50	2	4	3	1	4	3	3
12.60	2	4	9	3	4	2	1
11.80	2	4	6	2	4	1	2
11.20	3	2	2	1	3	2	2
12.10	3	2	9	3	4	2	1
9.70	3	2	10	4	1	2	3
7.80	3	3	3	1	4	3	3
9.60	3	3	11	4	2	3	2
10.60	3	3	4	2	1	3	1
8.70	3	1	12	4	3	1	1
9.10	3	1	8	3	2	1	3
8.00	3	1	6	2	4	1	2
8.70	3	4	1	1	2	4	1
7.50	3	4	7	3	1	4	2
8.30	3	4	5	2	3	4	3

'his analysis uses the data and gives an approximate table of means

**** ANALYSIS OF VARIANCE *****

'ARIATE: Y

OURCE OF VARIATION	DF	SS	SS%	MS	VR
STRATUM	2	16.1072	21.75	8.0536	8.329
.B STRATUM					
T	9	36.9025	49.84	4.1003	
OTAL	9	36.9025	49.84	4.1003	4.240
.B.*UNITS* STRATUM					
T	11	8.4630	11.43	0.7694	0.796
RESIDUAL	13	12.5703	16.98	0.9669	
OTAL	24	21.0333	28.41	0.8764	
RAND TOTAL	35	74.0430	100.00		

RAND MEAN 10.21
)TAL NUMBER OF OBSERVATIONS 36

**** INFORMATION SUMMARY *****

)DEL TERM EF NON-ORTHOGONAL TERMS

.B STRATUM
T1 0.407
T2 0.323 T1
T3 0.099 T1 T2

```
R.B.*UNITS* STRATUM
  T1                          0.593  R.B
  T2                          0.556  R.B  T1
  T3                          0.494  R.B  T1  T2

ALIASED MODEL TERMS
T


***** TABLES OF MEANS *****

VARIATE: Y


    GRAND MEAN     10.21

*** FOLLOWING TABLE OF MEANS IS ONLY APPROXIMATE ***

   T       1       2       3       4       5       6       7       8       9      10      11

  T1       1       1       1       2       2       2       3       3       3       4       4
  T2       2       3       4       1       3       4       1       2       4       1       2
  T3       4       2       3       3       4       1       4       1       2       2       3
  T4       1       2       3       1       3       2       2       3       1       3       2
       11.24   10.03    9.25    9.47    9.98    9.42   10.18   11.29   10.69    9.75   10.60

   T      12

  T1       4
  T2       3
  T3       1
  T4       1
      10.66

  ***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

  TABLE                    T
  ------------------------------
  REP                      3
  SED                  1.160
  EXCEPT WHEN COMPARING MEANS WITH SAME LEVEL(S) OF:
    T1                 0.991
    T2                 0.979
    T3                 0.954
    T4                 1.088


  ***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

  STRATUM                  DF          SE         CV%

  R                         2       0.819         8.0
  R.B                       0           *           *
  R.B.*UNITS*              13       0.983         9.6


'UNITS'          UNIT $ 108 = 1...108
'VARIATE'        NY = (Y)3
'FACTOR'         NTP $ 4 = T1,T2,T3
    :            COPY  $ 3 = 36(1...3)
    :            NT4 $ 3 = (T4)3
    :            DUMMY $ 36 = (1...36)3
    :            NT $ 12 = (T)3
    :            NR $ 3 = (R)3
    :            NB $ 4 = (B)3
'DEVALUE'        Y,T1,T2,T3,T,R,B
'DESCRIBE'       NY $ 3
'LINES'          10
'TREATMENT'      NT//(NTP*COPY+NT4)
'BLOCKS'         NR/NB/DUMMY
'CAPTION'        ''
This analysis uses copies of the data and in the analysis of variance table
gives three times the numbers in the above analysis ''
'ANOVA/PR=03013'  NY
'RUN'
```

```
This analysis uses copies of the data and in the analysis of variance table
gives three times the numbers in the above analysis
1
***** ANALYSIS OF VARIANCE *****

VARIATE: NY

SOURCE OF VARIATION                 DF          SS      SS%          MS        VR

NR STRATUM                           2   4.832E  1    21.75   2.416E  1

NR.NB STRATUM
  NT                                 9   1.107E  2    49.84   1.230E  1
TOTAL                                9   1.107E  2    49.84   1.230E  1

NR.NB.DUMMY STRATUM
  NT                                11   2.539E  1    11.43   2.308E  0     0.796
  RESIDUAL                          13   3.771E  1    16.98   2.901E  0
TOTAL                               24   6.310E  1    28.41   2.629E  0

NR.NB.DUMMY.*UNITS* STRATUM
  NT                                11   0.000E  0     0.00   0.000E  0
  RESIDUAL                          61   0.000E  0     0.00   0.000E  0
TOTAL                               72   0.000E  0     0.00   0.000E  0

GRAND TOTAL                        107   2.221E  2   100.00

GRAND MEAN                          10.214
TOTAL NUMBER OF OBSERVATIONS           108


***** INFORMATION SUMMARY *****

MODEL TERM                          EF  NON-ORTHOGONAL TERMS

NR.NB STRATUM
  NTP                             0.012
  NTP.COPY                        0.198

NR.NB.DUMMY STRATUM
  NTP                             0.099  NR.NB
  NTP.COPY                        0.247  NR.NB

NR.NB.DUMMY.*UNITS* STRATUM
  NTP                             0.889  NR.NB  NR.NB.DUMMY
  NTP.COPY                        0.556  NR.NB  NR.NB.DUMMY

ALIASED MODEL TERMS
NT


***** TABLES OF EFFECTS *****

VARIATE: NY


*** NR.NB STRATUM ***

 NTP                          EFFECTS:                REP        27  ESE

        NTP         1         2         3         4
             -7.625     3.575    -1.025     5.075


 NTP.COPY                     EFFECTS:                REP         9  ESE

       COPY         1         2         3
        NTP
          1     0.969    -1.356     0.388
          2    -1.456    -1.531     2.988
          3    -1.881     1.494     0.388
          4     2.369     1.394    -3.762
```

```
** NR.NB.DUMMY STRATUM ***

 NTP                          EFFECTS:                REP         27  ESE     1.0430

          NTP          1          2          3          4
                  -0.725      0.800      0.062     -0.137


 NTP.COPY                     EFFECTS:                REP          9  ESE     1.1425

         COPY          1          2          3
          NTP
            1     -0.160     -0.640      0.800
            2     -1.055      1.570     -0.515
            3      1.025     -0.145     -0.880
            4      0.190     -0.785      0.595

* NR.NB.DUMMY.*UNITS* STRATUM ***

 NTP                          EFFECTS:                REP         27  ESE     0.0000

          NTP          1          2          3          4
                   0.000      0.000      0.000      0.000

 COPY                         EFFECTS:                REP         36  ESE     0.0000

         COPY          1          2          3
                   0.000      0.000      0.000

 TP.COPY                      EFFECTS:                REP          9  ESE     0.0000

         COPY          1          2          3
          NTP
            1      0.000      0.000      0.000
            2      0.000      0.000      0.000
            3      0.000      0.000      0.000
            4      0.000      0.000      0.000

***** TABLES OF MEANS *****

VARIATE: NY

    GRAND MEAN    10.214

          NT          1          2          3          4          5          6          7          8          9         10         11

         NTP          2          2          3          2          3          2          3          2          3          2          3
        COPY          2          2          2          2          2          2          2          2          2          2          2
         NT4          1          2          3          1          3          2          2          3          1          3          2
              10.517     10.058     10.067     10.517     10.067     10.058     10.058     10.067     10.517     10.067     10.058


          NT         12

         NTP          3
        COPY          2
         NT4          1
              10.517
```

***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE	NT
REP	9
SED	0.4014

***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM	DF	SE	CV%
NR	2	0.8192	8.0
NR.NB	0	*	*
NR.NB.DUMMY	13	0.9833	9.6
NR.NB.DUMMY.*UNITS*	61	0.0000	0.0

'CLOSE'

R Thompson
Agricultural Research Council
Unit of Statistics
University of Edinburgh
James Clerk Maxwell Building
The King's Buildings
Mayfield Road
Edinburgh EH9 3JZ
U.K.

A REGRESSION MODEL FOR GENOTYPICAL STABILITY

INTRODUCTION

Yates and Cochran (1938) have described a model for relating the performance of a set of genotypes across a set of environments. The model has been used in studies of crop variety performance (Finlay and Wilkinson, 1963, Eberhart and Russell, 1966) and also in the evaluation of variety-isolate interactions in studies of disease resistance and pathogen virulence (Leonard and Moll, 1979, Jenns and Leonard, 1981).

The model takes the form

$$Y_{ij} = g_i + b_i I_j + d_{ij}$$

where Y_{ij} is response of the ith genotype grown in the jth environment,
g_i is the mean of the ith genotype,
b_i is the coefficient of regression for the response of the ith genotype on I_j, an environmental index, and
d_{ij} is deviation from regression.

The environmental index is variously taken as the mean response of all genotypes within an environment or as the mean response of a specific genotype within an environment. Generally, the coefficients, b_i, are used to characterise stability, or general adaptation, of the set of genotypes over the series of environments that are covered by the test. For further details on interpretation of the model, the reader is referred to the literature.

13

The use of multiple copies of data in forming and interpreting analysis of variance.

by

R. Thompson

Chapter 11

THE USE OF MULTIPLE COPIES OF DATA IN FORMING AND INTERPRETING ANALYSIS OF VARIANCE

Robin Thompson
University of Edinburgh
Edinburgh, Scotland

I. INTRODUCTION

In this paper the use of multiple copies of data is illustrated as an aid in both forming and interpreting analyses of variance. My interest in this technique arose from trying to analyse diallel crosses using the computer program GENSTAT (Alvey, Galwey and Lane, 1982). This program includes a general algorithm for the analyses of data with a high degree of balance and orthogonality (Wilkinson, 1970; Payne and Wilkinson, 1977). Whilst Wilkinson's algorithm is very powerful and is said to be able to analyse all the designs in Cochran and Cox (1957) (Heiberger, 1981) one cannot even specify the models appropriate for diallel crosses. Whilst one could use regression techniques to analyse such designs, these ignore the symmetry of the data and so can be computationally inefficient and less easy to interpret.

It was discovered that if the data was extended by taking two copies then this general algorithm can still be used. One associates different fixed effects with each copy and introduces another random

factor. Then the terms in the diallel cross analysis of variance correspond to terms in the analysis of variance of this extended set. In Section II this example is explained.

In field experiments the use of blocking to reduce the effect of soil heterogeneity has been used for many years. An alternative approach (P) originally suggested by Papadakis (1937) is based on adjusting yields by covariance on the residuals of neighbouring plots. Bartlett (1938) stressed the theoretical complications and the method was little used for some time, partly because of the tedious calculation (Yates, 1970).

Interest in the method has been stimulated by two recent discussion papers read to the Royal Statistical Society. In the first Bartlett (1978) extended Atkinson's (1969) results to show that the Papadakis treatment estimates are close to the maximum likelihood estimates under an autoregressive model. Usually solutions of the maximum likelihood equations for the variance parameters have to be found iteratively and this suggests estimating the Papadakis covariate regression coefficient iteratively using the hopefully improved estimates of residuals in each iteration.

In the second discussion paper Wilkinson, Hancock, Eckert and Mayo (1983) found that iterating the Papadakis method led to biased tests of treatment effects. The average adjusted treatment mean square was not equal to the expectation of the residual mean square for simulations using several sets of uniformity data. This led Wilkinson et al. to introduce an alternative procedure (NN) based on using 'sliding' blocks rather than fixed blocks. They used for each plot a block centered on each plot and including neighbours of that plot. This idea is appealing because with fixed blocks weights given to treatment comparisons from adjacent plots depend on whether or not the adjacent plots are deemed to be in the same block. In some circumstances one would want to give the same weight to both comparisons.

Wilkinson et al. (1983) derived estimators using intuitive, if not altogether consistent, arguments based on the classical analogue of recovery of inter-block information from incomplete blocks (Yates,

1936a). They produced evidence, again from uniformity trials, that suggested that tests of their treatment estimates were less biased. However, it was pointed out in the discussion that the NN treatment estimating equations were in fact equivalent to the P estimating equations, and the difference in the methods lay in the method of estimating the variance parameters.

This leads to the question of whether there are any other links between neighbour analysis and incomplete block analysis. There are at least two reasons why this is important. Firstly, there is instinctive prejudice against adjustment by covariates based on residuals (Kempton and Howes, 1981) suggesting that this and incomplete block methods are completely different, yet the link uncovered by the discussants of the Wilkinson et al. paper suggests common ground underlying the methods. Secondly, there are presumably circumstances when neighbour analysis, giving equal weight to treatment comparisons from adjacent plots is more efficient than incomplete block analysis. Then one would like to allocate treatments to plots in an optimum manner. If one can uncover more links between neighbour analyses and incomplete block analyses then perhaps optimum incomplete block designs, of which there is a huge literature, can easily be modified to give efficient designs for neighbour analysis.

These ideas are developed in Section III and again the technique of multiple copies of data is found useful using blocks of two superimposed on the data.

It is perhaps appropriate to explain why this paper should be in a volume dedicated to Professor Kempthorne. Firstly, all the classical designs mentioned in this paper are in his books on experimental design and genetic statistics. Secondly, several years ago after I had attended a talk advocating iteratively reweighted least squares as the solution to every statistical problem Kemp pointed out to me that many sets of genetic data, perhaps the simplest being ABO blood group data, did not fit into this framework. Later Bob Baker and I (1981) introduced the idea of composite link functions that allowed these genetic models to be embedded into the iteratively

reweighted least squares framework. Composite link functions link an expanded expectation of observation vector to the observation vector. The technique in this paper is closely related for here both observations and their expectation are expanded.

II. APPLICATION TO DIALLEL CROSS MODELS AND OTHER SIMILAR MODELS

In this section Wilkinson's (1970) algorithm for analysis of experimental designs is briefly discussed. Diallel cross models are introduced and it is shown how these can be fitted into the algorithm. The same procedure can be used on other designs and examples are given.

Wilkinson's Algorithm

Wilkinson's algorithm is a recursive procedure for the analysis of experimental designs. It involves a finite sequence of sweeps, in each of which a set of effects for a factor are calculated and an updated residual vector formed. The calculated effects are usually means calculated from the current residual vector, or 'effective' means which are means divided by an efficiency factor (Yates, 1936a). It is obvious that the procedure can be used for orthogonal experiments, with only two sweeps, one for blocks and one for treatments required for randomized blocks. Wilkinson uses a preliminary analysis of special dummy variates to calculate efficiency factors and eliminate redundant sweeps. The algorithm can deal with stratified designs having more than one error component by modifying the input observations for each stratum.

Wilkinson in his description of the algorithm allows a factor to have k associated efficiency factors and then k sweeps are carried out each with a different efficiency. Because of difficulties in calculating degrees of freedom associated with these k efficiencies the GENSTAT implementation of the algorithm only allows each factor to have one efficiency factor. Pseudo-factors can often be introduced, to allow GENSTAT analyses. Sometimes the necessary pseudo-

factors have been used in generating the design, as in lattices (Yates, 1936b).

To use the algorithm in GENSTAT the linear model to be fitted is specified in terms of qualitative fixed factors (called "Treatments" in GENSTAT), qualitative random factors ("Blocks") and quantitative variates ("Covariates").

Diallel Crosses

In plant breeding work on p different parental lines often all reciprocal crosses can be made and the p^2 progeny used for comparison of parental lines. Various models, analyses and interpretations have been given for data with this structure (Kempthorne, 1956; Griffing, 1956; Jinks and Mather, 1971). My limited interest here is in showing how analyses can be constructed using Wilkinson's algorithm and start by using a simple model that exhibits the difficulty in fitting these classes of models. Suppose that when a female of line i is crossed with a male of line j there is a progeny with observation y_{ij}. A factorial model $y_{ij} = \mu + f_i + m_j + e_{ij}$, where μ, f_i, m_j and e_{ij} are respectively mean, female, male and error terms. Sometimes, on genetic or other arguments, it can be argued that the male and female effects are equivalent and then a model

$$y_{ij} = \mu + \ell_i + \ell_j + e_{ij} \qquad (1)$$

where ℓ_i represents a line effect, is appropriate. Now each cross ($i \neq j$) observation includes two line terms and pure ($i=j$) plants includes the same term twice. Hence the model (1) cannot be specified directly using qualitative factors in GENSTAT terms.

Diallel crosses are not the only models that have more than one level of a factor related to an "observation". Yates (1936a) in his discussion of incomplete block experiments suggested an inter-block analysis as a first stage in making use of information from blocks. This analysis can be regarded as fitting a linear model to block totals and of course the expectation of the totals will include sums

of individual treatment effects. As Wilkinson's algorithm can do analyses in several strata this suggests extending the diallel cross data and constructing a model so that an inter-block analysis corresponds to model (1). This can be done by taking two copies of the data with $m_{ijI} = m_{ijII} = y_{ij}$ and using the models $m_{ijI} = \mu + \ell_i + u_{ij} + e_{ijI}$, $m_{ijII} = \mu + \ell_j + u_{ij} + e_{ijII}$, where u_{ij} is a random unit effect and e_{ijI} and e_{ijII} are random unit × copy effects. Female line effects are associated with the first copy and male effects with the second copy. Then an inter-unit analysis of m_{ij} gives multiples of the required line effects and sums of squares. The line effects have efficiency $(p-2)/(p-1)$ as noted by Wilkinson (1970).

Other models and designs can be fitted using the same technique. Hayman (1954) introduced a model that includes a comparison of mean (and line) pure performance with mean (and line) cross performance and also line effects on reciprocal differences $(y_{ij} - y_{ji})$. The extended model for the two copies of the data are

$$m_{ijI} = \mu + \ell_i + p_k + \ell p_{ik} + c_I + \ell.c_{iI} + r_s + r.h_{st} + e_{ijI}$$

$$m_{ijII} = \mu + \ell_j + p_k + \ell p_{jk} + c_{II} + \ell.c_{jII} + r_s + r.h_{st} + e_{ikII}$$

where p_k is a fixed effect with two levels representing pure and cross ($k=1$ if $i=j$, $k=2$ if $k \neq j$) and h_t is a fixed effect with two levels representing halves of the table ($t=1$ if $i<j$, $t=2$ if $i \geq j$), c_i a fixed effect representing copies of the data ($i=I$, II), r_s is a random factor with $p \times (p+1)/2$ levels representing reciprocals with crosses i,j and j,i at level $i \times (i-1)/2+j$. Some interactions of line, copy, reciprocal and half are also in the model. This gives the following skeleton analysis of variance:

Source	Degrees of freedom
Reciprocal Stratum	
Lines	p-1
Pure versus Cross	1
Lines × Pure versus Cross	p-1
Residual	p(p-3)/2
Total	p(p+1)/2-1

Half within Reciprocal Stratum

Lines × Copy	$p-1$
Residual	$(p-1)(p-2)/2$
Total	$p(p-1)/2$

Within Half within Reciprocal Stratum

Lines	$p-1$
Copy	1
Lines × Copy	p
Residual	$(p-1)^2$
Total	p^2

The sums of squares in the lowest stratum are zero because they are comparisons between the two copies and the other sums of squares are four times (because two copies of the data are used) those in Hayman's analysis.

Other Designs

Elsewhere (1983) the author has applied the same idea to other designs. I make use of the fact that a pseudo-factorial model similar to (1) can be imposed on partially balanced incomplete blocks with triangular structure (Clatworthy, 1973). With a little more difficulty rectangular lattices (Harshbarger, 1947) can be dealt with in a similar manner.

III. INCOMPLETE BLOCK AND NEIGHBOUR ANALYSIS

In this section links between Incomplete Block and Neighbour analysis are explored. A multiple copy model is introduced to interpret terms arising in an autoregressive model. This assists in the explanation of terms in the NN and P methods, and allows the consideration of efficiency factors and helps in constructing designs. The contributions of various strata to variance estimates and extensions to more appropriate models are discussed.

One Dimensional Auto-Regressive Model

Suppose there are $n=rt$ plots, comprising r for each of t treatments, laid out in one long strip and that the ith plot receives treatment j and a model

$$y_i = \tau_j + e_i \quad (i=1,\ldots,n;\ j=1,\ldots,t) \tag{2}$$

is appropriate, where e_i is a plot error with mean zero, variance σ^2 and covariance with e_k of $\rho^{|i-k|}\sigma^2$. This autoregressive variance structure has been used to explicate the P method by Atkinson (1969) and Bartlett (1978). This model (2) cannot be defended in all circumstances. For example, the plots normally should be arranged in replicates and separate replicate effects often need to be included. This layout in replicates or other experimental constraints in the field will sometimes imply that the autoregressive structure applies to groups of less than rt plots. However, this model (2) will be used because it illustrates several important points that carry over to other, more appropriate and complicated, models.

It is convenient to write (2) in matrix form

$$E(y) = X\tau,\ \mathrm{var}(e) = \sigma^2 H \tag{3}$$

where X is a n×t matrix with elements zero and one and H is a n×n matrix with elements $H_{ij} = \rho^{|i-j|}$. Then if ρ and σ^2 were known the least squares estimates of the treatment effects, τ_A, satisfy

$$(X'H^{-1}X)\tau = X'H^{-1}y \text{ with } \mathrm{var}(\tau_A) = \sigma^2(X'H^{-1}X)^{-1} \tag{4}$$

If matrices D and L, of size n×n, are defined with all elements of D and L zero except $D_{11} = D_{nn} = 1$ and $L_{i,i+1} = L_{i+1,i} = 1$ then H^{-1} can be written as

$$(1-\rho^2)H^{-1} = [(1+\rho^2)(I-D) + D-\rho L] \tag{5}$$

Alternatively if $2B = 2I-D+L$ and $2W = 2I-D-L$ then

$$2(1-\rho^2)H^{-1} = (1-\rho)^2B + (1+\rho)^2W + (1-\rho^2)D \tag{6}$$

$$(1-\rho^2)H^{-1} = (1+\rho)^2[I - 1/2\ D + (1-\rho)/2(1+\rho)D - 2\rho/(1+\rho)^2B] \tag{7}$$

By comparison a mixed model often used with incomplete block experiments is $y = X\tau + Zb + e$ where Z represents allocation of plots to blocks and b represents block effects with variance Γ and e are independent plot errors with variance σ_e^2. Using both intra- and inter-block information this leads to estimating equations similar to (5) with H^{-1} now given by

$$(I - Z(Z'Z + I\Gamma^{-1})Z') \tag{8}$$

and σ_e^2 replacing σ^2. If Γ^{-1} is put equal to zero in (8) then the intra-block estimating equations result.

Comparison of (7) and (8) suggests considering a conceptual model. Suppose two copies of the data are taken, with $m_I = m_{II} = y$, and two sets of blocks of two are superimposed. In the first copy (I) plots 2i and 2i+1 are in the ith I block and in the second copy (II) plots 2i-1 and 2i are in the ith II block (i=1,...,n). Diagramatically

I copy Block	1		2		3		4		5	.....
Plot	1	2	3	4	5	6	7	8		.....
II copy Block	1		2		3		4		5	.....

Then consider the incomplete block model similar to the one above with

$$m_{Ii} = \tau_j + b_{Ir} + e_{Ii},\ m_{IIi} = \tau_j + b_{IIs} + e_{IIi}$$

or (9)

$$m_I = X\tau + Z_I b_I + e_I,\ m_{II} = X\tau + Z_{II} b_{II} + e_{II}$$

where r and s are integer parts of (i+1)/2 and (i+2)/2 respectively. Plots 1 and n will each be in blocks of size one once and without loss of generality suppose treatment 1(2) is applied to plot 1(n).

If the block effects are independent and size two (one) block effects have variance $\gamma_2\sigma_e^2(\gamma_1\sigma_e^2)$ then $Z(Z'Z + \Gamma^{-1})Z'$ for this model is $1/(1+\gamma_1^{-1})D + 2/(2+\gamma_2^{-1})B$. So if $\gamma_1 = 2\rho/(1-\rho)$, $\gamma_2 = 2\rho/(1-\rho)^2$ and $\sigma_e^2 = 2\sigma^2(1-\rho^2)/(1+\rho^2)$ the parameters of the incomplete block model can be chosen to mimic the autoregressive model. Equations (6) and (9) suggests that τ_A combines information from between blocks of two (B), within blocks of two (W) and from the end plots (D), the weighting depending on ρ.

Model (9) can also be used to interpret terms arising in the estimation of σ^2 and ρ by maximizing the likelihood of error contrasts (REML) (Patterson and Thompson, 1971) if e in (5) is normally distributed. Terms arise in sums of squares of residuals $r_c' r_c$ and $\hat{b}_c' \hat{D}_c \hat{b}_c$ where $\hat{b}_c = (Z_c'Z_c + \Gamma_c^{-1})^{-1} Z_c'm_c$, $r_c = m_c - X\tau - Z_c b_c$ are residuals and $D_c = \Gamma_c^{-1}(d\Gamma_c/d\rho)\Gamma_c^{-1}$ (c = I,II). These terms also arise if (9) were the appropriate model. In both models these sums of squares of residuals are equated to their expectation. The expectations under (5) and (9) differ but there are similar terms in both models associated with corrections for τ being estimated.

Connection with P and NN Methods

An approximation, suggested by (5), is to use V^{-1} proportional to $(\Delta-\beta L)$ where $\Delta = I - 1/2\ D$ and $\beta = \rho/(1+\rho^2)$ helps to explain the P and NN methods. Then (4) becomes

$$X'(\Delta-\beta L)X\tau = X'(\Delta-\beta L)y \qquad (10)$$

or

$$X'\Delta X\tau = X'(\Delta y-\beta L(y-X\tau)) \qquad (11)$$

As $L(y-X\tau)$ has ith element the sum of residuals from the (i-1) and (i+1) plots then (10) can be thought of as using this vector as a covariate and this is one justification for the P method.

Wilkinson et al. (1983) consider using 'sliding' blocks that is for plot i considering a block of plots (i-1), i and (i+1), which

also leads, apart from differential weighting to border plots, to (11). In particular they discuss intra-(sliding) block treatment estimates with $2\beta=1$ which corresponds to replacing V^{-1} by W in (5) and is equivalent to the intra-block treatment estimates from the multiple copy model.

The methods differ on variance estimation. The P method uses analogues with the analysis of covariance to estimate β although no correction is made for the estimation of τ in the covariate $L(y-X\tau)$. Both P and NN use sums of squares of residuals of the form $R = (y-X\tau)'(I-\beta L)^2(y-X\tau)$. The use of R arises naturally and the residuals can be related to the sum of residuals in the multiple copy model, $(r_I + r_{II})$. R will be an efficient statistic under some variance models, for example $V^{-1} \propto (I-\beta L)^2$, but it is not an efficient statistic when the treatment estimates are efficient (i.e., $V^{-1} \propto (I-\beta L)$). The implicit use of these two variance models in the P method partly explains the problems in estimating β iteratively.

Efficiency Factors

For incomplete block models it is often useful to consider contrasts that are independently estimated. Their efficiency factors play a fundamental role in the structure of the design (James and Wilkinson, 1971; Pearce, Caliński and Marshall, 1974). The intrablock estimate of τ from (9) satisfies $(2R-X'BX)\tau_w = X'(2I-D-B)$ where $R = X'(I - 1/2\ D)X$ is a diagonal matrix. Because $I - (2R)^{-1/2}X'BX(2R)^{-1/2} = A$ is a symmetric matrix there exists an orthogonal matrix such that $A = P\varepsilon P'$, where ε is a diagonal matrix containing the eigenvalues of A. The columns of P are eigenvectors of A and the columns can be ordered according to the eigenvalues. If the design is connected there will be only one zero eigenvalue corresponding to an eigenvector $(2R)^{1/2}/(2n)^{1/2}$. The matrix $C = (2R)^{1/2}P$ can be constructed and it can be shown (Pearce et al., 1974) that the first t-1 elements of $C'\tau_w$ represent contrasts of τ_w. The variance of $C'\tau_w$ is $\sigma_e^2\varepsilon^-$, ε^- is a diagonal matrix with elements ε_i^{-1} (or zero if $\varepsilon_i = 0$). So $C'\tau_w$ represent independent contrasts. The variance of

$C'(2R)^{-1}X'(2I-D)y$ is $I\sigma_e^2$ if $\Gamma = 0$ so that ε_i can be thought of as an efficiency factor for the i-th contrast.

Similarly if a combined estimate τ_R is found using the intra- and inter-block (of size two) information then for model (9) the variance of $C'\tau_c$ is $\sigma_e^2(1+2\gamma_2)[I+2\gamma_2\varepsilon]^{-1}$ so that $\varepsilon_{Ri} = (1+2\gamma_2\varepsilon_i)/(1+2\gamma_2)$ is an efficiency factor for $C'\tau_i$.

The variance matrix for $C'\tau_A$ using the autoregressive model is slightly more complicated due to the contribution of end plots. Then using (7) we find that

$$(1-\rho^2)X'H^{-1}X = (1+\rho^2)R^{1/2}[(1-\rho^2)/(1+\rho^2)I + 4\rho/(1+\rho)^2A$$

$$+ (1-\rho)/2(1+\rho)R^{-1/2}X'DXR^{-1/2}]R^{1/2}$$

Then it can be shown that the i,j element of the variance matrix of τ_A is

$$2(1-\rho^2)\sigma^2/(1+\rho)^2[\varepsilon_{Ri}^{-1}\delta_{ij} - \varepsilon_{Ri}^{-1}\varepsilon_{Rj}^{-1}\sum_{m,n=1}^{2} P_{mi}P_{mj}F_{mn}] \qquad (12)$$

where F and G are 2×2 matrices with $F = G^{-1}$ and elements of G given by $G_{mn} = \sum_i P_{mi}\varepsilon_{Ri}^{-1}P_{ni} + \delta_{mn}(2r-1)(1+\rho)/(1-\rho)$, $\varepsilon_{Ri} = (1+2\gamma_2\varepsilon_i)/(1+2\gamma_2) = [(1-\rho)^2 + 4\rho\varepsilon_i]/(1+\rho)^2$ and δ_{ij} is the Kronecker delta. The first term in G_{mn} relates to the covariance of τ_{Rm} and τ_{Rn}. The second term in (12) represents contributions to the covariance from the end plots, this term should be negligible for most contrasts, and reduce as ρ increases, especially if the end treatments, 1 and 2, do not contribute much to the contrast. The second term can contribute substantially to the variance of $2R/(2n)^{1/2}\tau_A$, especially as ρ increases, because $\varepsilon_t = 0$. When $\rho = 0$ then (12) reduces to $2\sigma^2\{1-(P_{1i}P_{1j} + P_{2i}P_{2j})/2r\}$. This suggests the contrasts are approximately uncorrelated and that $\varepsilon_{Ai} = [(1-\rho)^2 + 4\rho\varepsilon_i]/(1-\rho^2)$ is an efficiency factor.

To quantify the gain from using two copies of the data consider the case when the auto-regressive model is appropriate but an

incomplete block model is used with blocks of two. Then $\sigma_e^2 = (1-\rho)\sigma^2$ and σ_b^2 is approximately $\rho\sigma^2$. This ignores the average covariance between plots not in the same block, of the order of $1/[2t(1-\rho)]$ if ρ is not too large. For a contrast with intra-block efficiency ε_i then using the block information increases the efficiency (relative to σ_e^2) to $[(1-\rho)+2\rho\varepsilon_i]/(1+\rho)$ and relative to σ^2 is $[(1-\rho)+2\rho\varepsilon_i]/(1-\rho^2)$ so the gain from using all the information is of the order of $\rho(2\varepsilon_i-1+\varepsilon)/(1-\rho^2)$. Note that the average efficiency from a 2r replicate design can be higher than from an r replicate design because there is more scope for balancing comparisons within blocks.

Variances for NN Method

Wilkinson et al. (1983) used a variance matrix given approximately, neglecting some border information, by

$$(I-\beta L)V(I-\beta L) = \sigma^2[(I-\beta L)(I-\beta L) + \Omega(I+L+L^2-2I)]$$

where $\Omega = (1-2\beta)^2 w$ and w and β are variance parameters, to give variances of their treatment estimates. Then

$$[(I-\beta L)(I-\beta L)+\Omega(I+L+L^2-2I)]$$

$$= (1+2\beta^2+\Omega)I + (\Omega-2\beta)(2B-2I) + (\Omega+\beta^2)(2B_1-2I)$$

where $2B_1 = 2I+(L^2-2I)$ and can be thought of as block information when 2 copies of the data and blocks of two are formed from plots one apart i.e., plots j and j+2. Suppose ε_1 a diagonal matrix exists such that $X(2R-B_1)X = P\varepsilon_1 P'$ that is the estimated contrasts $C'\tau$ would be independent with both blocking schemes. Then

$$X'(I-\beta L)X = rP[(1-2\beta)I + 4\beta\varepsilon]P'$$

and $X'(I-\beta L)V(I-\beta L) = \sigma^2 rP[((1-2\beta)^2 + \Omega)I + 8\beta\varepsilon - 4\beta\varepsilon_1 - 4\Omega(I-\varepsilon-\varepsilon_1)]P'$ so that an efficiency factor for the ith contrast is approximately $\varepsilon_{iNN} = [(1-2\beta)+4\beta\varepsilon_i]^2/[(1-2\beta)^2 + \Omega + 8\beta\varepsilon_i - 4\beta\varepsilon_{1i} - 4\Omega(1-\varepsilon_i-\varepsilon_{1i})]$.

Wilkinson et al. (1983) give a numerical example of a Rothamsted experiment (Jenkyn et al., 1979). This is an experiment balanced for first and second neighbours with efficiencies $e_i = 2/3$, $e_{1i} = 4$. Wilkinson et al. give efficiencies of 1.22 for an intra-block analysis which is approximately (3/2) ε_{iNN} using $2\beta=1$, $\Omega=0$. Wilkinson et al. use the factor 3/2 to scale their variances.

These efficiencies also arise in expressions for the bias in the P covariate.

Variance Parameter Estimation

The subdivision of the treatments into independent contrasts helps in subdividing the residual sums of squares $R = (y-X\tau_A)'H^{-1}(y-X\tau_A)$. By relating τ_A to τ_R and then relating τ_R to the estimates from $Wy(\tau_W)$ and $By(\tau_B)$ then it can be shown that

$$R = [(1-\rho)^2/(1+\rho)^2](R_W+T_{RD}) + R_B + T_{WB}$$

where R_W and R_B are residual sums of squares $y'Wy - \tau_W'C'\varepsilon C\tau_W$ and $y'By - \tau_B'C'(I-\varepsilon)C\tau_B$ and T_{WB} is a sum of squares for treatment comparisons between strata W and B, i.e.,

$$T_{WB} = (\tau_W-\tau_B)'C\ \varepsilon\ \varepsilon_R^-(I-\varepsilon)C'(\tau_W-\tau_B)$$

and similarly

$$T_{RD} = (2r-1)(y_1-\tau_{R1},\ y_n-\tau_{R2})F \begin{pmatrix} y_1 - \tau_{R1} \\ y_n - \tau_{R2} \end{pmatrix}$$

is a comparison between τ_R and estimates from the end plots. Yates (1940) when recovering inter-block information interpreted his analysis of variance in a similar way.

In REML estimation of σ^2 and ρ can be thought of as equating R and $dR/d\rho$ to their expectation. The contributions of R_W, R_B and T_{BW} to $\hat{\sigma}^2$ and $\hat{\rho}$ can be fairly easily calculated showing, for instance

that the weight given to T_{BW}, relative to R_B, increases as ρ increases. The contribution of T_{RD} is harder to quantify as it depends on $dF/d\rho$ but in some numerical examples T_{RD} seems to become more influential than T_{BW} as ρ increases. I am a little uneasy about this, given that T_{RD} is a comparison of border plots with others and border plots might behave slightly differently just because they are on the boundary.

Implications for Design

Efficient balanced designs are available for this variance model and for a limited number of treatment and replicate combinations (Williams, 1952; Freeman, 1979). The efficiencies of a derived 2r replicate incomplete block model play a key role in quantifying the efficiencies of an r replicate design with an autoregressive model. Much work has been done on constructing efficient incomplete block designs. These facts suggest that these 2r incomplete blocks designs might be converted into good r replicate designs.

For example, John, Wolock and David (1972) suggest a cyclic design with t = 18, k = 2 and r = 6 constructed cyclically from initial blocks of 1 2, 1 4 and 1 9. This suggests a 3 replicates autoregressive design.

1	2	5	6	9	10	13	14	17	18	3	4	5	8	11	12	15	16
3	6	16	13	5	8	18	15	7	10	2	17	9	12	4	1	11	14
4	5	15	14	6	7	17	16	8	9	1	18	10	11	3	2	12	13

where the rows represent replicates and where there is a constant difference between the ith and (i+4)th elements (i=1,...,14) in each row. Taking two copies of this design and imposing two sets of blocks of two as in (10) gives blocks included in the 6 replicate design.

Extension to Other Variance Structures

The autoregressive variance model was introduced to illustrate the use of a multiple copy model and relate it to other results but it will not necessarily be appropriate. Patterson and Hunter (1983)

from an analysis of 166 cereal trials in the United Kingdom suggested that an error model with $V_{ii} = (1+\theta)\sigma^2$, $V_{ij} = \rho^{|i-j|}\sigma^2$ might be more appropriate. It is interesting to show how this model fits into the multiple copy framework.

For this alternative model

$$V = \sigma^2\theta I + \sigma^2(1-\rho^2)[(1+\rho^2)(I-D) + D-\rho L]^{-1}$$

$$= \sigma^2\theta I + \sigma^2[(1-\rho)/(1+\rho)]V_A^{-1}$$

so that

$$V^{-1} \propto V_A - V_A[V_A+[(1-\rho)/(1+\rho)\theta]I]^{-1} V_A$$

The least squares equations for τ then satisfy

$$\begin{bmatrix} X'V_A X & X'V_A \\ V_A X & V_A+[(1-\rho)/(1+\rho)\theta]I \end{bmatrix} \begin{bmatrix} \tau \\ u \end{bmatrix} = \begin{bmatrix} X'V_A y \\ V_A y \end{bmatrix} \qquad (13)$$

This is rather like (4) and suggests taking 2 copies of the data and imposing blocks of two on the data as in model (9). Then (13) can be manipulated to give

$$\begin{bmatrix} 2X'X & X'Z_I & X'Z_{II} & 2X' \\ Z_I'X & Z_I'Z_I+\Gamma^{-1} & 0 & Z_I' \\ Z_{II}'X & 0 & Z_{II}'Z_{II}+\Gamma^{-1} & Z_{II}' \\ 2X & Z_I & Z_{II} & 2[1+(1-\rho)/(1+\rho)\theta]I \end{bmatrix} \begin{bmatrix} \tau \\ b_I \\ b_{II} \\ u \end{bmatrix} = \begin{bmatrix} 2X'y \\ Z_I' y \\ Z_{II}' y \\ 2y \end{bmatrix}$$

This suggests that a multiple copy model to mimic this error model is

$$m_I = X\tau + Z_I b_I + u + e_{Ii}$$
$$m_{II} = X\tau + Z_{II} b_{II} + u + e_{IIi} \qquad (14)$$

where u has variance $2\{\theta(1+\rho)\sigma^2/(1-\rho)\}I$. Hence there are two copies of the data, blocks of two superimposed and a unit effect u_i associated with each copy of plot i. The parameter θ in a sense measures the closeness of the two copies and if $\theta = 0$ or $-2(1-\rho)/(1+\rho)$ then auto-regressive or moving average models result. Although (14) is a conceptual model one can speculate whether the parameters have any physical interpretation. The layout suggests that if m_I and m_{II} were observations on plots harvested in halves then perhaps θ, ρ and σ^2 could be interpreted in terms of these observations.

Often designs are laid out in two dimensions, and again one can introduce more copies and blocks of two to mimic the structure.

Randomization

There remains the question of whether randomization of the multiple copies gives added justification for the use of specific variance models. Certainly randomizations are possible that mimic the structure of V^{-1}.

IV. CONCLUSION

It has been shown that multiple copies give added insight into two specific examples. The idea can be applied in other areas. Space only allows me to say that problems of symmetry in x and y can sometimes be fruitfully expressed in terms of x, y and y, x and that the use of unit, copy and unit × copy information can be used to investigate extra variation in exponential family models (Hinde, 1982).

ACKNOWLEDGMENTS

I am grateful for encouragement from Dr. H.D. Patterson and other members of the ARC Unit.

REFERENCES

Alvey, N., Galwey, N., and Lane, P. (1982). *An Introduction to Genstat*. London: Academic Press.

Atkinson, A.C. (1969). The use of residuals as a concomitant variable. *Biometrika 56*, 33-41.

Bartlett, M.S. (1938). The approximate recovery of information from field experiments with large blocks. *J. Agric. Sci. 28*, 418-427.

Bartlett, M.S. (1978). Nearest neighbour models in the analysis of field experiments. *J. Roy. Stat. Soc. B 40*, 147-174.

Clatworthy, W.H. (1973). *Tables of two-associate-class partially balanced designs. Nat. Bur. Standards Appl. Math. Ser.* 63.

Cochran, W.G. and Cox, G.M. (1957). *Experimental Designs*. New York: John Wiley.

Freeman, G.H. (1979). Some two-dimensional designs, balanced for nearest neighbours. *J. Roy. Stat. Soc. B 41*, 88-95.

Griffing, B. (1956). A generalized treatment of the use of diallel crosses in quantitative inheritance. *Heredity 10*, 31-50.

Harshbarger, B. (1947). Rectangular Lattices. Virginia Agricultural Experimental Station Memoir 1.

Hayman, B.I. (1954). The analysis of variance of diallel tables. *Biometrics 10*, 235-244.

Heiberger, R.M. (1981). The specification of experimental designs to ANOVA programs. *American Statistician 35*, 98-104.

Hinde, J. (1982). *Compound Poisson Regression Models. GLIM82*, R. Gilchrist (ed.). New York: Springer.

James, A.T. and Wilkinson, G.N. (1971). Factorisation of the residual operator and canonical decomposition of nonorthogonal factors in the analysis of variance. *Biometrika 58*, 279-294.

Jenkyn, J.F., Bainbridge, A., Dyke, G.V. and Todd, A.D. (1979). An investigation into inter-plot interactions, in experiments with mildew on experiments with mildew on barley, using balanced designs. *Ann. Appl. Biol. 92*, 11-28.

John, J.A., Wolock, F.W. and David, H.A. (1972). *Cyclic designs. Nat. Bur. Standards Appl. Math.* Ser., 62.

Kempthorne, O. (1956). The theory of the diallel cross. *Genetics 41,* 451-459.

Kempton, R.A. and Howes, C.W. (1981). The use of neighbouring plot values in the analysis of variety trials. *Applied Statistics 30,* 59-70.

Mather, K. and Jinks, J.L. (1971). *Biometrical Genetics.* Ithaca: Cornell Univ. Press.

Papadakis, J.S. (1937). Méthode statistique pour des expériences sur champ. *Bull Inst. Amél. Plantes á Salonique 23.*

Patterson, H.D. and Hunter, E.A. (1983). The efficiency of incomplete block designs in national list and recommended list cereal variety trials. *J. Agric. Sci.* (to be published).

Patterson, H.D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika 58,* 545-554.

Payne, R.W. and Wilkinson, G.N. (1977). A general algorithm for analysis of variance. *J. Roy. Stat. Soc.* Ser. C *26,* 251-260.

Pearce, S.C., Calinski, T. and Marshall, T.F. de C. (1974). The basic contrasts of an experimental design with special reference to the analysis of data. *Biometrika 61,* 449-460.

Thompson, R. (1983). Diallel Cross, Incomplete Block designs and rectangular Lattices. *Genstat Newsletter 9* (in press).

Thompson, R. and Baker, R.J. (1981). Composite link functions in generalized linear models. *Appl. Statistics 30,* 125-131.

Wilkinson, G.N. (1970). A general recursive procedure for analysis of variance. *Biometrika 57,* 19-46.

Wilkinson, G.N., Eckert, S.R., Hancock, T.N. and Mayo, O. (1983). Nearest neighbour (NN) analysis of field experiments. *J. Roy. Stat. Soc. B* (in press).

Williams, R.M. (1952). Experimental designs for serially correlated observations. *Biometrika 39,* 151-167.

Yates, F. (1936a). Incomplete randomised blocks. *Ann. Eugen. 7,* 121-140.

Yates, F. (1936b). A new method of arranging variety trials involving a large number of varieties. *J. Agric. Sci 26,* 424-455.

Yates, F. (1940). The recovery of inter-block information in balanced incomplete block designs. *Ann. Eugen. 10,* 223-230.

Yates, F. (1970). *Experimental Design.* London: Griffin.

14

The estimation of maternal genetic variances

by

R. Thompson

BIOMETRICS *32*, 903–917
DECEMBER 1976

The Estimation of Maternal Genetic Variances

ROBIN THOMPSON

Statistical Laboratory, Iowa State University, Ames, Iowa, U.S.A.
and
ARC Unit of Statistics, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, Scotland

Summary

The estimation of maternal genetic variances by a multivariate maximum likelihood method is discussed. As an illustration the method is applied to data on Tribolium using a model based on partitioning the maternal genetic effect into additive and dominance components. An alternative model due to Falconer [1965] is also fitted. The method is applied to designs suggested for estimating maternal variances by Eisen [1967]. Modifications required when parents are selected on their phenotypic values are given.

1. Introduction

Classically the variance of quantitative traits is partitioned into genotypic components due to the genotype of the individual and environmental components. When this partition is not sufficient, indirect effects can be hypothesized. In mammals maternal effects are important indirect effects. The effect of the mother influences the environment of her offspring but can have a genetic component. Just as the genetic component due to the individual can be partitioned into additive, dominance and epistatic components (Fisher [1918]), so can the maternal genetic component (Willham [1963]).

These maternal effects are important to animal breeders who would like to eliminate the influence of maternal effects so that selection is for direct genetic merit. Interest also exists in maternal effects for improving maternal performance.

The usual method of estimating maternal genetic variances is to calculate covariances between near relatives over two generations and to equate these to their expected values in terms of the genetic and environmental variances. Sometimes, in well-designed experiments, there are more covariances between relatives than parameters and a least-squares procedure has been suggested, for example by Van Vleck and Hart [1966] and Eisen [1967], as a way of combining the information available. This procedure is not fully efficient if the variances and covariances of the covariances between relatives are not homogeneous. It is possible to use a weighted least squares procedure, but this involves the calculation of the covariances of the covariances between relatives which while possible is very tedious.

In this paper a maximum likelihood (ML) approach, based on the multivariate analysis of variance, is considered. In Section 2 we outline the theory needed. In Section 3 the method is illustrated with some data on *Tribolium castaneum* given by Bondari [1971] and we fit an alternative model due to Falconer [1965]. In section 4 we show how the ML method

Journal Paper No. J-9261 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa. Project 1669.

Key Words: Maternal genetic variances; Maximum likelihood estimation; Design of genetic experiments.

can be applied to designs suggested by Eisen [1967] for the estimation of maternal genetic variances. Section 5 shows how the method can be modified when parents are selected on their phenotypic values. For consistency with previous literature a large number of symbols have had to be used. These are listed in Appendix 1.

2. *Theory*

The key idea is to subdivide the data into independent parts according to different sources of variation, as in the multivariate analysis of variance. For each line of the analysis of variance we have a symmetric sum of squares and products matrix representing variation in a number of variates. The only difference from the usual multivariate analysis of variance is that the numbers of variates associated with different sources of variation need not be equal.

A simple example is a hierarchical design with each of a set of sires mated to several dams and a number of offspring raised from each mating (Thompson [1976], Hill and Nicholas [1974]). An analysis of variance can easily be constructed for the offspring data with separate sums of squares within dams, between dams within sires, and between sires. This basic structure can be extended to include the data on the parents by forming two matrices: a 2×2 matrix, representing the variation between dams within sires with two variates corresponding to offspring and dams, and a 3×3 matrix, representing variation between sires with three variates corresponding to offspring, dams and sires.

In general we have s symmetric sums of squares and product matrices $\mathbf{S}_k$ ($k = 1, \cdots, s$) of size s_k and these are independent of each other. If d_k are the degrees of freedom associated with $\mathbf{S}_k$, a mean square and product matrix $\mathbf{M}_k$ is given by $\mathbf{S}_k/d_k$. Let $\mathbf{V}_k$ be the expected value of $\mathbf{M}_k$. We suppose that it can be written as

$$\mathbf{V}_k = \sum_{i=1}^{p} \mathbf{F}_{ki} \cdot \theta_i \tag{1}$$

where θ_i ($i = 1, \cdots, p$) are the variance parameters of interest and $\mathbf{F}_{ki}$ are known ($s_k \times s_k$) symmetric matrices. We assume that the observations are normally distributed and estimate the θ's by maximizing, not the likelihood of all the data, but that of the $\mathbf{S}$'s. The advantage of such an approach is discussed by Patterson and Thompson [1971]. The log likelihood, L, can be written as

$$L = \text{constant} - \frac{1}{2} \sum_{k=1}^{s} d_k(\ln |\mathbf{V}_k| + \text{tr}\,(\mathbf{M}_k\mathbf{V}_k^{-1})) \tag{2}$$

where $|\mathbf{V}|$ represents the determinant of $\mathbf{V}$ and tr ($\mathbf{V}$) the trace (sum of diagonal elements).

The maximum likelihood estimates of θ_i satisfy (Anderson [1973])

$$2\delta L/\delta\theta_i = \sum_{k=1}^{s} d_k \,\text{tr}\,(\mathbf{V}_k^{-1}\mathbf{M}_k\mathbf{V}_k^{-1}\mathbf{F}_{ki}) - \sum_{k=1}^{s} d_k \,\text{tr}\,(\mathbf{V}_k^{-1}\mathbf{F}_{ki}) = 0 \tag{3}$$

for $i = 1, \cdots, p$ or $\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$ where $\mathbf{A}$ is a $p \times p$ matrix with i, jth element

$$A_{ij} = \sum_{k=1}^{s} d_k \,\text{tr}\,(\mathbf{V}_k^{-1}\mathbf{F}_{ki}\mathbf{V}_k^{-1}\mathbf{F}_{kj}), \tag{4}$$

$\mathbf{b}$ is a $p \times 1$ column vector with ith element

$$b_i = \sum_{k=1}^{s} d_k \,\text{tr}\,(\mathbf{V}_k^{-1}\mathbf{M}_k\mathbf{V}_k^{-1}\mathbf{F}_{ki}), \tag{5}$$

and $\boldsymbol{\theta}$ is a $p \times 1$ column vector with ith element θ_i.

It is usually necessary to solve (3) iteratively. A suitable scheme is to get initial estimates of $\boldsymbol{\theta}$, evaluate $\mathbf{V}_k$, $\mathbf{A}$ and $\mathbf{B}$ (equations (1), (3) and (4)); solve for $\boldsymbol{\theta}$, and then repeat using this new estimate of $\boldsymbol{\theta}$. Alternatively in the first iteration, $\mathbf{M}_k$ could be used as an initial estimate of $\mathbf{V}_k$ in the evaluation of $\mathbf{A}$ and $\mathbf{B}$. The asymptotic variance-covariance matrix of $\boldsymbol{\theta}$ is given by $2\mathbf{A}^{-1}$.

If a linear combination of $\mathbf{F}_{ki}$, say $\Sigma l_i \mathbf{F}_{ki}$, is equal to the zero matrix for all k, then the matrix $\mathbf{A}$ is singular, which means that not all the parameters can be estimated. In this case, as in ordinary linear model theory, we can impose constraints on the parameters or reparameterize the model in order to make progress.

The procedure outlined here is equivalent to the weighted least squares method mentioned in the introduction but is much easier to implement. For example, to use the weighted least squares method for one of Eisen's [1967] designs would require the evaluation of 91 variances and covariances of covariances between relatives.

The ML method is probably most useful when the data are balanced. When the data are unbalanced full ML methods can be computationally unfeasible. However estimation methods for variance components with balanced data have, by analogy, suggested methods for unbalanced data (Searle [1971]). Similarly consideration of the form of the procedure outlined in this paper may suggest reasonably efficient methods for unbalanced data. The formulae developed for balanced data are also useful for comparing different designs and give approximate bounds on the precision of estimates from unbalanced data.

3. *Example*

Design and Model.

In this section data presented by Bondari [1971] are used to illustrate how the theory outlined in the previous section can be applied to the estimation of genetic maternal effects. Bondari used two mating designs, suggested by Willham [1963], to investigate genetic maternal influences on pupa weight and family size in *Tribolium castaneum*. In both designs grandsires were mated to random granddams and two offspring, one male and one female, were measured from each mating. These first generation offspring were allowed to mate and yield second generation progeny according to mating plans. In design (a) two paternal half-sibs of different sexes were each mated to a random mate. In design (b) two female paternal half-sibs were each mated to a random mate. Two progeny (one of each sex) of each mating were measured. The mating plans can be obtained from Figure 1.

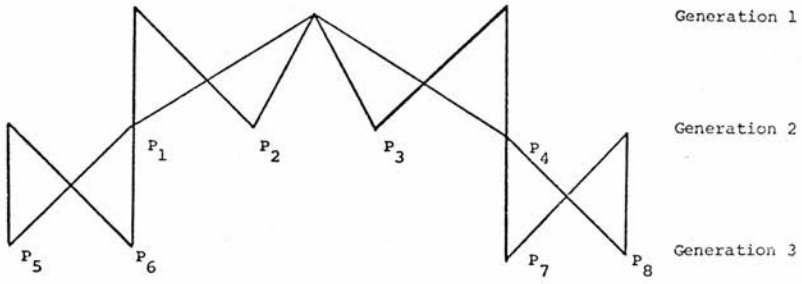


Figure 1

MATING PLAN FOR BONDARI'S DESIGNS

In design (a) P_1 and P_3 are males, P_2 and P_4 are females.
In design (b) P_1 and P_4 are females, P_2 and P_3 are males.

In Table 1 the types of relatives possible in generations 2 and 3 from the two designs are indicated. If we assume a random mating population in linkage equilibrium, a maternal genetic effect and the absence of epistasis then the covariances between relatives (Y_j) can be written in terms of nine parameters given in Table 1.

Three terms, $\sigma^2_{A_o}$, $\sigma^2_{D_o}$ and $\sigma^2_{E_o}$, represent the variances that arise in the classical subdivision of variance into additive, dominance and environmental components. Three terms, $\sigma^2_{A_m}$, $\sigma^2_{D_m}$ and $\sigma^2_{E_m}$, represent the same partitioning for the hypothesized maternal effects and three terms, $\sigma_{A_oA_m}$, $\sigma_{D_oD_m}$ and $\sigma_{E_oE_m}$, represent the corresponding covariances between direct and maternal effects.

To simplify some of the results we find it convenient to replace the mnenomic double suffix form of the parameters. We therefore let θ_i be the variance parameter in the ith column in Table 1; for example $\theta_5 = \sigma^2_{A_m}$.

The covariance between relatives, Y_j, can then be written as $\sum_{i=1}^{9} T_{ji}\theta_i$ with coefficients T_{ji} given in Table 1. Thus, for example, the covariance between dam and offspring (Y_9) is

$$\tfrac{1}{2}\theta_1 + \tfrac{5}{4}\theta_3 + 1\theta_4 + \tfrac{1}{2}\theta_5 + 1\theta_9 = \tfrac{1}{2}\sigma_{A_o}{}^2 + \tfrac{5}{4}\sigma_{A_oA_M} + \sigma_{D_oD_M} + \tfrac{1}{2}\sigma_{A_M}{}^2 + \sigma_{E_oE_M}.$$

Formulae for the T_{ji} are given by Willham [1963].

Application of Method.

Bondari measured pupae weight for 208 families (i.e., group of eight individuals $P_1, \cdots, P_8$ in Figure 1) from design (a) and from 123 families from design (b). We define

Table 1

THE COMPOSITION OF COVARIANCES BETWEEN RELATIVES GENERATED BY FIVE DESIGNS. THE ELEMENTS ARE THE COEFFICIENTS (T_{ji}) OF THE COMPONENTS GIVEN AS COLUMN HEADS. * INDICATES THAT A PARTICULAR RELATIONSHIP OCCURS IN A GIVEN DESIGN. (*a*) AND (*b*) REPRESENT BONDARI'S DESIGNS (SECTION 3) AND *I*, *II* AND *III* REPRESENT EISEN'S DESIGNS (SECTION 4).

Relationship	Causal components of variance and covariance										Mating Design				
	i	1	2	3	4	5	6	7	8	9					
	j	$\sigma^2_{A_O}$	$\sigma^2_{D_O}$	$\sigma_{A_OA_M}$	$\sigma_{D_OD_M}$	$\sigma^2_{A_M}$	$\sigma^2_{D_M}$	$\sigma^2_{E_M}$	$\sigma^2_{E_O}$	$\sigma_{E_OE_M}$	(a)	(b)	I	II	III
Paternal half sibs	1	1/4	0	0	0	0	0	0	0	0	*	*	*	*	*
Single first cousins (sires full sibs) or Paternal half uncle (or aunt) (paternal half sibs) with nephew (or niece)	2	1/8	0	0	0	0	0	0	0	0	*		*	*	*
Paternal half sibs plus single first cousins (dams full sibs)	3	3/8	1/8	1/2	0	1/2	1/4	0	0	0			*	*	*
Three-quarter sibs (dams paternal half sibs)	4	5/16	1/16	1/4	0	1/4	0	0	0	0			*		*
Double first cousins (sires full sibs and dams full sibs)	5	1/4	1/16	1/2	0	1/2	1/4	0	0	0			*	*	*
Single first cousins (sires full sibs) plus half first cousins (dams paternal half sibs)	6	3/16	1/32	1/4	0	1/4	0	0	0	0			*		*
Full sibs	7	1/2	1/4	1	0	1	1	1	0	0	*	*	*	*	*
Within full sibs	8	1/2	3/4	0	0	0	0	0	1	0	*	*	*	*	*
Dam-offspring	9	1/2	0	5/4	1	1/2	0	0	0	1	*	*	*	*	*
Sire-offspring	10	1/2	0	1/4	0	0	0	0	0	0	*		*	*	*
Maternal uncle (or aunt) with nephew (or niece)	11	1/4	0	3/4	1/4	1/2	0	0	0	0		*	*	*	*
Maternal half uncle (or aunt) (paternal half sibs) with nephew (or niece) or Single first cousin (opposite sexes full sibs)	12	1/8	0	1/4	0	0	0	0	0	0	*	*	*	*	*
Paternal uncle (or aunt) with nephew (or niece)	13	1/4	0	1/4	0	0	0	0	0	0	*		*	*	*
Double first cousins (opposite sexes full sibs)	14	1/4	1/16	1/2	0	0	0	0	0	0				*	*
Single first cousins (dams full sibs)	15	1/8	0	1/2	0	1/2	1/4	0	0	0				*	
Half first cousins (dams paternal half sibs)	16	1/16	0	1/4	0	1/4	0	0	0	0		*			*
Half first cousins (opposite sexes half sibs)	17	1/16	0	1/8	0	0	0	0	0	0	*				

two 8 × 8 matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ of the variances and covariances between families within the two designs. We let y_{ika} be the measurement on P_i from the kth family from design (a) and let $y_{ioa} = \sum_{k=1}^{208} y_{ika}$. The i, jth element of $\mathbf{S}_1$, S_{1ij} is then $\sum_{k=1}^{208} y_{ika}y_{jka} - (y_{ioa}y_{joa})/208$. The degrees of freedom associated with $\mathbf{S}_1$, d_1, are 207. Similarly, we define $\mathbf{S}_2$ and d_2 using data from design (b). Mean squares and products matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ are calculated by dividing $\mathbf{S}_1$ and $\mathbf{S}_2$ by d_1 and d_2. These matrices are symmetric; the upper triangle of $\mathbf{M}_1(M_{1ij}\ ;\ j \geq i)$ and the lower triangle of $\mathbf{M}_2(M_{2ij}\ ;\ i \geq j)$ are given in Table 2 for the variate (pupae weight/200) measured in micrograms. We use $\mathbf{M}_1$ and $\mathbf{M}_2$ to estimate the variance parameters. Some contrasts between y_{ioa} and y_{iob} could provide extra information on the variance parameters but the extra information is so small (about 0.2 percent) compared with that from $\mathbf{M}_1$ and $\mathbf{M}_2$ that it can be neglected.

In Table 3 are given the upper triangle of $\mathbf{V}_1$ and the lower triangle of $\mathbf{V}_2$ in terms of the covariances between relatives Y_j. Since Y_j can be written as $\sum_{i=1}^{9} T_{ji}\theta_i$, $\mathbf{F}_{ki}$ can be found by replacing Y_j by T_{ji} in the expression for $\mathbf{V}_k$. For example, the matrix below contains the upper triangle of $\mathbf{F}_{11}$ (diagonal and above) and the lower triangle of $\mathbf{F}_{21}$ (diagonal and below)

$$\begin{pmatrix}
1 & 1/2 & 1/4 & 1/4 & 1/2 & 1/2 & 1/8 & 1/8 \\
1/2 & 1 & 1/4 & 1/4 & 1/4 & 1/4 & 1/8 & 1/8 \\
1/4 & 1/4 & 1 & 1/2 & 1/8 & 1/8 & 1/4 & 1/4 \\
1/4 & 1/4 & 1/2 & 1 & 1/8 & 1/8 & 1/2 & 1/2 \\
1/2 & 1/4 & 1/8 & 1/8 & 1 & 1/2 & 1/16 & 1/16 \\
1/2 & 1/4 & 1/8 & 1/8 & 1/2 & 1 & 1/16 & 1/16 \\
1/8 & 1/2 & 1/4 & 1/2 & 1/16 & 1/16 & 1 & 1/2 \\
1/8 & 1/2 & 1/4 & 1/2 & 1/16 & 1/16 & 1/2 & 1
\end{pmatrix}$$

Table 2

UPPER TRIANGLE OF MEAN SQUARE AND PRODUCT MATRIX FOR DESIGN (*a*). $\mathbf{M}_1$, (ABOVE DIAGONAL) AND LOWER TRIANGLE OF MEAN SQUARE AND PRODUCT MATRIX FOR DESIGN (*b*), $\mathbf{M}_2$, (BELOW DIAGONAL) FOR PUPAE WEIGHT/200 (μg).

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	
	0.8419	0.3793	0.0038	0.0608	0.1164	0.0276	0.0430	0.0719	P_1
P_1	0.7973	0.9886	0.0505	0.0200	0.0954	0.0297	0.0205	0.0123	P_2
P_2	0.2906	0.8957	0.6260	0.2492	0.0617	0.1514	0.1618	0.1324	P_3
P_3	0.1931	0.2257	0.9968	0.8599	0.0307	0.0211	0.2607	0.2340	P_4
P_4	0.1097	0.1895	0.3767	1.0208	0.7672	0.3417	0.0097	0.0256	P_5
P_5	0.3140	0.1367	0.2277	-0.0120	0.7698	0.9651	0.0104	0.0084	P_6
P_6	0.2769	0.1034	0.1943	0.0587	0.3180	0.6357	0.7841	0.3361	P_7
P_7	0.2150	0.1064	0.1813	0.2138	0.0398	0.1749	0.8653	0.8694	P_8
P_8	0.1483	0.1221	0.0880	0.3874	-0.0613	0.0899	0.2917	0.8565	
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	

The **F** matrices are related as follows:

$$\mathbf{F}_{k6} = \mathbf{F}_{k7} \quad \text{and} \quad 8\mathbf{F}_{k2} = \mathbf{F}_{k6} + \mathbf{F}_{k7} + 6\mathbf{F}_{k8} \quad (k = 1, 2).$$

This means we cannot estimate all the parameters. However, we can estimate θ_1 , θ_3 , θ_4 , θ_5 , θ_9 , θ_{10} and θ_{11} where $\theta_{10} = \frac{1}{4}\theta_2 + \theta_6 + \theta_7$ and θ_{11} is the phenotypic variance σ^2_P $(= \theta_1 + \theta_2 + \theta_3 + \theta_5 + \theta_6 + \theta_7 + \theta_8)$.

Table 3

EXPECTED VALUE OF M_1 , $\mathbf{V}_1$, (DIAGONAL AND ABOVE) AND EXPECTED VALUE OF $\mathbf{M}_2$, $\mathbf{V}_2$, (DIAGONAL AND BELOW) IN TERMS OF Y_i , THE COVARIANCES BETWEEN RELATIVES

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
P_1	$Y_8 + Y_7$	Y_7	Y_1	Y_1	Y_{10}	Y_{10}	Y_{12}	Y_{12}
P_2	Y_7	$Y_8 + Y_7$	Y_1	Y_1	Y_{13}	Y_{13}	Y_{12}	Y_{12}
P_3	Y_1	Y_1	$Y_8 + Y_7$	Y_7	Y_2	Y_2	Y_{11}	Y_{11}
P_4	Y_1	Y_1	Y_7	$Y_8 + Y_7$	Y_2	Y_2	Y_9	Y_9
P_5	Y_9	Y_{11}	Y_{12}	Y_{12}	$Y_8 + Y_7$	Y_7	Y_{17}	Y_{17}
P_6	Y_9	Y_{11}	Y_{12}	Y_{12}	Y_7	$Y_8 + Y_7$	Y_{17}	Y_{17}
P_7	Y_{12}	Y_{12}	Y_{11}	Y_9	Y_{16}	Y_{16}	$Y_8 + Y_7$	Y_7
P_8	Y_{12}	Y_{12}	Y_{11}	Y_9	Y_{16}	Y_{16}	Y_7	$Y_8 + Y_7$

If we express the covariances between relatives, Y_i , in terms of these parameters then the **F** matrices for these seven parameters can be found by replacing Y_i in Table 3 in turn by $T_{i1} - T_{i8}$, $T_{i3} - T_{i8}$, $T_{i4} - T_{i8}$, $T_{i5} - T_{i8}$, $T_{i9} - T_{i8}$, $T_{i6} - T_{i8}$ and T_{i8} . The solution to the set of equations (3) was found by the iterative scheme outlined in Section 2 with initial estimates of one for the phenotypic variance, and zero for the other parameters. In Table 4 we give the resulting estimates and the asymptotic variance-covariance matrix. The estimated values of $\mathbf{M}_1$ (diagonal and above) and $\mathbf{M}_2$ (diagonal and below) using these estimates are

$$\begin{bmatrix}
0.8426 & 0.3218 & 0.0483 & 0.0483 & 0.1205 & 0.1205 & 0.0481 & 0.0481 \\
0.3218 & 0.8426 & 0.0483 & 0.0483 & 0.0722 & 0.0722 & 0.0481 & 0.0481 \\
0.0483 & 0.0483 & 0.8426 & 0.3218 & 0.0241 & 0.0241 & 0.1470 & 0.1470 \\
0.0483 & 0.0483 & 0.3218 & 0.8426 & 0.0241 & 0.0241 & 0.2709 & 0.2709 \\
0.2709 & 0.1470 & 0.0481 & 0.0481 & 0.8426 & 0.3218 & 0.0240 & 0.0240 \\
0.2709 & 0.1470 & 0.0481 & 0.0481 & 0.3218 & 0.8426 & 0.0240 & 0.0240 \\
0.0481 & 0.0481 & 0.1470 & 0.2709 & 0.0276 & 0.0276 & 0.8426 & 0.3218 \\
0.0481 & 0.0481 & 0.1470 & 0.2709 & 0.0276 & 0.0276 & 0.3218 & 0.8426
\end{bmatrix}.$$

Table 4

ESTIMATES OF VARIANCE AND COVARIANCE COMPONENTS WITH ASYMPTOTIC VARIANCE-COVARIANCE MATRIX (COVARIANCES ABOVE DIAGONAL, CORRELATIONS BELOW).

$\hat{\sigma}^2_P$ =	$\hat{\sigma}^2_{A_O}$ =	$\hat{\sigma}_{A_O A_M}$ =	$\hat{\sigma}^2_{A_M}$ =	$\hat{\sigma}_{D_O D_M}$ =	$\hat{\theta}_{1O}$* =	$\hat{\sigma}_{E_O E_M}$ =
0.0843	0.193	0.096	-0.033	0.175	0.163	-0.103
0.0007	0.0004	0.0002	0.0000	0.0002	0.0001	-0.0002
0.17	0.0073	-0.0035	-0.0009	0.0024	-0.0007	-0.0018
0.08	-0.44	0.0089	-0.0025	-0.0176	-0.0045	0.0096
-0.00	0.06	-0.15	0.0324	-0.0574	-0.0304	0.0441
0.02	0.07	-0.44	-0.76	0.1768	0.0742	-0.1267
0.02	-0.04	-0.25	-0.90	0.94	0.0355	-0.0532
-0.02	-0.07	0.33	0.80	-0.99	-0.92	0.0935

* $\theta_{1O} = \frac{1}{4}\sigma^2_{D_O} + \sigma^2_{D_M} + \sigma^2_{E_M}$

One of the difficulties of this model is that the coefficients of some of the components for the covariances are very similar between the classes of relatives. This suggests that the parameters may be imprecisely estimated. Table 5 shows that for the Tribolium data the estimates of $\sigma_{D_O D_M}$, $\frac{1}{4}\sigma^2_{D_O} + \sigma^2_{D_M} + \sigma^2_{E_M}$ and $\sigma_{E_O E_M}$ are relatively imprecise and highly correlated. However, as in the case when direct effects alone are hypothesized, most, if not all, interest is in the functions of the additive and phenotypic components. The dominance and environmental components are usually of little interest and are often only thought of as necessary by-products of estimating the additive and phenotypic components. The most frequently derived and interpreted parameters are the heritability σ^2_A/σ^2_P and $(\sigma^2_{A_O} + (3/2)\sigma_{A_O A_M} + (1/2)\sigma^2_{A_M})/\sigma^2_P$, the fraction of the selection differential that is realized if selection is on phenotypic values (Dickerson [1947]). These have values 0.23 and 0.38 for the Tribolium data with approximate standard errors 0.10 and 0.17. The estimates are not very precise considering that over 2000 larvae were measured. The designs were suggested (Willham [1964]) for two reasons—because they are feasible for farm animals and also because $\sigma^2_{A_O}$, $\sigma^2_{A_M}$ and $\sigma_{A_O A_M}$ can easily be estimated from covariances Y_1, Y_{16} and Y_{17}—but he gave no consideration to the precision of the estimates. Some improvement in precision follows if one is willing to assume a priori that some of the parameters are zero. This may or may not be warranted; This author prefers to leave such a decision to the geneticist.

Falconer's Model.

Because of the low precision of, and high correlations between some of the estimates an alternative model, used by Falconer [1965], is now fitted. He suggested that a model for the phenotypic value of an individual, P, might be expressed as the sum of five terms—the individual's breeding value; mP' the maternal effect as a linear function m of the mother's phenotype; the individual's dominance deviation; the effect of environmental factors

Table 5

COEFFICIENTS T_{ji} ($i = 10, \cdots, 13$) OF COVARIANCES BETWEEN RELATIVES FROM DESIGNS (*a*) AND (*b*) FOR FALCONER'S MODEL AND ESTIMATED COVARIANCES

Coeff	T_{j10}	T_{j11}	T_{j12}	T_{j13}	$\hat{Y}_j$
Coeff of	σ^2_P	σ^2_A	$(m/2(2-m))\sigma^2_A$	σ^2_D	
j					
1	0	1/4	0	0	0.0576
2	0	1/8	0	0	0.0288
7	m^2	1/2	4	1/4	0.3194
8	$1-m^2$	-1/2	-4	-1/4	0.5220
9	m	1/2	1	0	0.2691
10	0	1/2	1	0	0.1260
11	m^3	(1+2m)/4	1+4m	m/4	0.1227
12	0	(1+2m)/8	0	0	0.0386
13	0	1/4	1	0	0.0684
16	0	$(1+4(m+m^2))/16$	0	0	0.0259
17	0	(1+2m)/16	0	0	0.0193

common to full-sibs that are not included in the maternal effect; all other environmental effects.

It is convenient to write the covariance between relatives as $Y_i = T_{i10}\sigma^2_P + (T_{i11} + T_{i12}(m/2(2-m))\sigma^2_A + T_{i13}\sigma^2_D$, where σ^2_P is the phenotypic variance, σ^2_A the additive variance and σ^2_D the dominance variance. In Table 5 are given the coefficients T_{ji} ($i = 10, \cdots, 13$) for the covariances arising from designs (a) and (b). For example, the covariance between full-sibs (Y_7) is $m^2\sigma^2_P + \sigma^2_A/2 + [4m/2(2-m)]\sigma^2_A + \sigma^2_D/4$. The negative coefficients in Y_8 arise because $Y_7 + Y_8$ is the total phenotypic variance.

Estimates of σ^2_P, σ^2_A and σ^2_D were found for different values of m since $\mathbf{V}_1$ and $\mathbf{V}_2$ are non-linear in m. The maximum of L, the log-likelihood, corresponded to $\hat{m} = 0.17$, $\hat{\sigma}^2_P = 0.842$, $\hat{\sigma}^2_A = 0.231$ and $\hat{\sigma}^2_D = 0.548$. The asymptotic variance-covariance matrix of the estimates $\hat{\sigma}^2_P$, $\hat{\sigma}^2_A$ and $\hat{\sigma}^2_D$ is

$$\begin{pmatrix} 0.0007 & 0.0005 & 0.0004 \\ 0.45 & 0.0020 & -0.0036 \\ 0.10 & -0.66 & 0.0164 \end{pmatrix}$$

with covariances above the diagonal and correlations below the diagonal. This is calculated assuming the true value of $m = 0.17$ and will underestimate the sampling variances. Estimates of the covariances between relatives are given in Table 5.

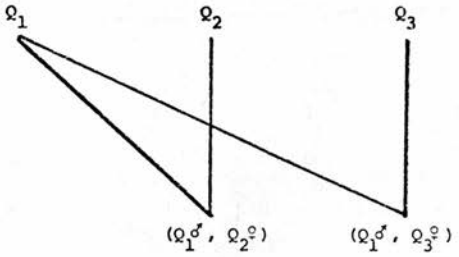
The estimation procedure has been described in terms of two 8×8 matrices. Alternatively one could work with two independent matrices for design (a). One, of size 1×1 gives differences between full-sibs and one of size 6×6 gives differences between families with rows and columns corresponding to P_1, P_2, P_3, P_4, $P_5 + P_6$ and $P_7 + P_8$, respec-

tively. In design (b) this 6 × 6 matrix can be split into two 3 × 3 independent matrices. The rows and columns correspond to offspring, female parents and male parents, respectively. One matrix represents differences within grandsire families (i.e., P_1 , P_2 , P_5 , P_6 , compared with P_3 , P_4 , P_7 and P_8) and one represents differences between grandsire families.

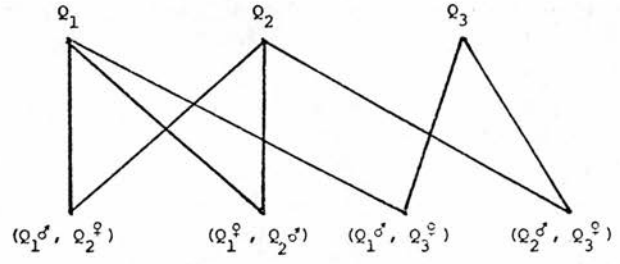
4. Eisen's Designs

Eisen [1967] suggested three designs that might be used for estimating direct and maternal variances and covariances. These designs are more suitable for laboratory animals than for farm animals. The three designs follow the same pattern in that from initial random matings in the first generation R sets are sampled, and three unrelated families Q_1 , Q_2 and Q_3 are contained in each set. In design I, Q_1 is a full-sib family of males, Q_2 is a full-sib family of females. In designs II and III Q_1 and Q_2 are full-sib families of males and females. In designs I and III Q_3 is a paternal half-sib family of females and in design II Q_3 is a full-sib family of females. The mating plans can be obtained from Figure 2. In design I each of the

I Q_1 *full-sib family (males),* Q_2 *full-sib family (females),* Q_3 *half-sib family (females)*



II Q_1 *and* Q_2 *full-sib families (males and females),* Q_3 *full-sib family (females)*



III Q_1 *and* Q_2 *full-sib families (males and females),* Q_3 *half-sib family (females)*

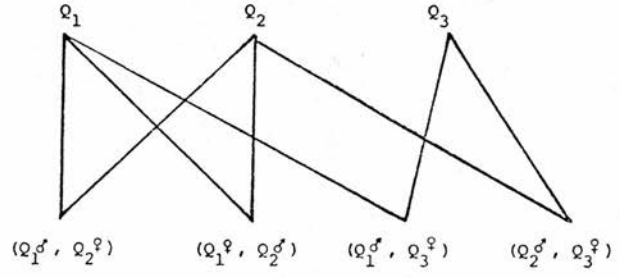


Figure 2

MATING PLANS FOR EISEN'S DESIGNS

S males from Q_1 is mated to D_1 females from Q_2 and to D_2 females from Q_3 . In designs II and III each of the S males from $Q_1(Q_2)$ is mated to D_1 females from $Q_2(Q_1)$ and to D_2 females from Q_3 . M offspring are observed from each mating. The types of relative possible in the second and third generations are given in Table 1 for the three designs.

The table is taken mainly from Eisen [1967] but designs II and III generate more covariances than Eisen indicated. The extra covariances occur only if more than one sire is used from the full-sib families Q_1 and Q_2 . We will assume, as Eisen assumed, that the direct-maternal environmental covariance $\sigma_{E_O E_M}$ is zero.

In Appendix 2 we show how the procedure outlined in Section 2 can be applied to these designs. Essentially the data are split into seven independent sums of squares and products matrices, and these form the basis of the analysis.

It is natural to ask if one design is better than another and what are the optimal values of the design parameters S, D_1 and D_2 . N, the litter size, will usually be fixed by the species that is considered and so there is little scope for varying it. The possible designs are to a large extent limited by the number of full-sibs in a family. Assuming an equal number of males and females in a litter, the maximum S is $N/2$ and the maximum D_1 is $N/2S$. There is more scope for the choice of D_2 , the number of dams from half-sib families in designs I and III, but in designs II the maximum D_2 is $N/4S$. Not all choices of S, D_1 and D_2 allow the estimation of all eight parameters. For example, $\sigma^2_{D_M} - \sigma^2_{E_M}$ and $\sigma^2_{D_O} - 2\sigma^2_{A_M} + 2\sigma^2_{D_M}$ are not estimable in design I, when $S = 1$ and $D_1 = 1$ and $4\sigma^2_{D_O} - 2\sigma^2_{D_M} + \sigma^2_{E_M}$ is not estimable when $S = 1$ and $D_2 = 1$. In design III $\sigma^2_{D_M} - \sigma^2_{E_M}$ is not estimable when $S = 1$ and $D_1 = 1$.

It is difficult to make an exhaustive comparison of the designs as up to eight parameters are estimated and the parameters vary in importance. The asymptotic variance-covariance matrix $(2\mathbf{A}^{-1})$ on which a comparison of designs will be based is dependent on the actual value of the parameters. However one or two comments seem in order, based on a study of the variance-covariance matrix for six sets of variance values. One set of values were suggested by the results on Tribolium discussed in Section 3, and another by the results of Hanrahan and Eisen [1974] on 12-day weight of mice. The four other sets were combinations of the first two.

Comparison of designs I and III, for fixed values of N, S, D_1 and D_2 , suggests that the variances/observation of estimates from design III are usually smaller than from design I. Consideration of the variances of heritability $(\sigma^2_{A_O}/\sigma^2_P)$ and of $((\sigma^2_{A_O} + (3/2)\sigma_{A_O A_M} + (1/2)\sigma^2_{A_M})/(\sigma^2_P)$ suggests that for design III it is reasonable to make S as large as possible, $D_1 = 1$ and D_2 so that $SD_2 = 8$. The comparisons between designs II and III are less clear cut. They suggest that, for the same value of N, design II is better for the estimation of dominance and environmental parameters and design III is better for additive parameters.

Estimates for traits with maternal effects are less precise than estimates on traits with only direct effects. For example, the variances of heritability/observation for the optimal designs when there are maternal effects are roughly three to five times the variances of heritability/observation for the optimal designs. (Hill and Nicholas [1974]) when there are only direct effects.

5. Parents Subject to Culling

Sometimes animals that are used as parents of the next generation are chosen on their phenotypic performance, for example, Van Vleck and Hart [1966]. In this section the modifications to the method of Section 2 to deal with this are outlined.

First suppose there are data on a parent generation, with sum of squares matrix $\mathbf{S}_1$ and expectation matrix $\mathbf{V}_1$, and some of these animals have offspring. The sum of squares and products matrix between parents and offspring $\mathbf{S}_2$ and expectation matrix $\mathbf{V}_2$ can be partitioned as

$$\mathbf{S}_2 = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix} \qquad \mathbf{V}_2 = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}$$

where the 1 suffix relates to the parent generation and the 2 to the offspring, and $\mathbf{V}_{11} \equiv V_1$. Then the log-likelihood becomes

$$L = \text{const} - \tfrac{1}{2}\{d_1 \ln |\mathbf{V}_1| + \text{tr}\,[(\mathbf{S}_1 - \mathbf{S}_{11})\mathbf{V}_1^{-1}] + d_2 \ln |\mathbf{V}_2| + \text{tr}\,\mathbf{S}_2\mathbf{V}_2^{-1}\}.$$

Then if $d_1\mathbf{M}_1 = \mathbf{S}_1 - \mathbf{S}_{11}$, equation (3) still holds even if parents are selected on their own performance. However the asymptotic variance-covariance matrix is not $2\mathbf{A}^{-1}$ but $2\mathbf{A}^{*-1}$ where the i, jth element of $\mathbf{A}^*$ is given by

$$A_{ij}^* = A_{ij} + 2\left\{\text{tr}\left[\begin{bmatrix} \mathbf{V}_1^{-1}(\mathbf{S}_{11} - d_1\mathbf{V}_1)\mathbf{V}_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}\mathbf{F}_{2i}\mathbf{V}_2^{-1}\mathbf{F}_{2j}\right]\right.$$
$$\left. - \text{tr}\,[\mathbf{V}_1^{-1}(\mathbf{S}_{11} - d_1\mathbf{V}_1)\mathbf{V}_1^{-1}\mathbf{F}_{1i}\mathbf{V}_1^{-1}\mathbf{F}_{1j}]\right\}. \qquad (6)$$

The iterative scheme in Section 2 can be used replacing $\mathbf{A}$ by $\mathbf{A}^*$.

Sometimes data are only available on the selected parents and the offspring. In this situation, the regression of offspring on parents can sometimes be used to estimate the heritability (Falconer [1960]). This can be justified by an argument similar to that of Henderson, *et al.* [1959] involving the conditional likelihood of the offspring's records given the parent's records. In our case the log-likelihood is

$$L = \text{const} - \tfrac{1}{2}d_2\,(\ln |\mathbf{V}_2| + \text{tr}\,(\mathbf{M}_2\mathbf{V}_2^{-1}) - \ln |\mathbf{V}_1| - \text{tr}\,(\mathbf{M}_{11}\mathbf{V}_1^{-1})).$$

Equations (3), (4) and (6) can be again used if d_1 is replaced by $-d_2$ and the iterative scheme outlined again seems useful. The difference between A_{ij}^* when all the parent generation is measured and when only the parents are measured is $(d_1 + d_2)$ tr $(\mathbf{V}_1^{-1}\mathbf{F}_{1i}\mathbf{V}_1^{-1}\mathbf{F}_{1j})$ which corresponds to the information on the $d_1 + d_2$ animals in the parent generation that is used in one case but not the other.

Acknowledgments

I am grateful to Oscar Kempthorne, R. L. Willham and W. G. Hill for encouragement and comments. Some of this work was done in connection with research under NIH Grant 13827 and under NSF Grant GP-24614.

L'Estimation des Variances Génétiques Maternelles

Résumé

On discute l'estimation des variances génétiques maternelles par une méthode multivariée de maximum de vraisemblance. A titre d'illustration, on applique la méthode à des données sur le Tribolium en utilisant un modèle fondé sur le découpage de l'effet génétique maternel en com-

posantes additives et de dominance. On ajuste également un autre modèle dû à Falconer [1965]. La méthode s'applique à des plans suggérés pour estimer les variances maternelles par Eisen [1967]. On indique les modifications nécessaires lorsque les parents sont sélectionnés selon leur valeur phénotypique.

References

Anderson, T. W. [1973]. Asymptotically efficient estimation of covariance matrices with linear structure. *Annals of Statistics 1*, 135–41.

Bondari, K. [1971]. A study of genetic maternal effects in a designed experiment using Tribolium. Unpublished Ph.D. thesis, Iowa State University Library, Ames, Iowa.

Dickerson, G. E. [1947]. Composition of hog carcasses as influenced by heritable differences in rate and economy of gain. *Iowa Agricultural Experimental Station Bulletin*, 489–524.

Eisen, E. J. [1967]. Mating designs for estimating direct and maternal genetic variances and direct-maternal covariances. *Canadian Journal of Genetic Cytology 9*, 13–22.

Falconer, D. S. [1960]. *Introduction to Quantitative Genetics.* Ronald Press Co., New York, New York.

Falconer, D. S. [1965]. Maternal effects and selection response. Genetics Today. S. J. Geerts, (ed.), Pergamon Press, New York, New York.

Fisher, R. A. [1918]. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Royal Soc., Edinburgh 52*, 399–433.

Hanrahan, J. P. and Eisen, E. J. [1974]. Genetic variation in litter size and 12-day weight in mice and their relationships with post-weaning growth. *Animal Prod. 19*, 13–22.

Henderson, C. R., Kempthorne, O., Searle, S. R. and Von Krosigk, C. M. [1959]. The estimation of environmental and genetic trends from records subject to culling. *Biometrics 15*, 192–218.

Hill, W. G. and Nicholas, F. W. [1974]. Estimation of heritability by both regression on parent and intra-class correlation of sibs in one experiment. *Biometrics 30*, 447–68.

Patterson, H. D. and Thompson, R. [1971]. Recovery of inter-block information when block sizes are unequal. *Biometrika 58*, 545–54.

Searle, S. R. [1971]. *Linear Models.* Wiley, New York.

Thompson, R. [1973]. The estimation of variance and covariance components with an application when records are subject to culling. *Biometrics 29*, 527–50.

Thompson, R. [1976]. Design of experiments to estimate heritability when observations are available on parents and offspring. *Biometrics 32*, 283–304.

Van Vleck, L. D. and Hart, C. L. [1966]. Covariances among first-lactation milk records of cousins. *Journal of Dairy Science. 49*, 41–5.

Willham, R. L. [1963]. The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics 19*, 18–27.

Willham, R. L. [1964]. The genetic relationship between milk production and growth rate: experimental approaches for evaluation. (Mimeo.) *Oklahoma Agricultural Station.*

Received November 1975, *Revised June* 1976

Appendix 1

In this appendix we give the main notation and indicate where the symbols are defined.

$\mathbf{A}$	$(p \times p)$ matrix used in estimation process	(4)
$\mathbf{A}^*$	modification to $\mathbf{A}$ when parents selected	(6)
$\mathbf{b}$	$(p \times 1)$ vector used in estimation process	(5)
D_i	$(i = 1, 2)$ number of females in Eisen's designs	(Section 4)
d_k	degrees of freedom associated with $\mathbf{S}_k$	(Section 2)
$\mathbf{F}_{ki}$	matrix of coefficients of θ_i in expected value of $\mathbf{M}_k$	(Section 2)
L	log-likelihood	(2)
$\mathbf{M}_k$	mean square and product matrix $(= \mathbf{S}_k/d_k)$	(Section 2)
m	coefficient of mother's phenotype in Falconer's model	(Section 2)

N	number of offspring from each mating	(Section 4)
P	phenotypic value on individual in Falconer's model	(Section 2)
P'	phenotypic value on mother in Falconer's model	(Section 2)
P_i	identification of individuals in Bondari's design	(Section 3)
Q_i	(i = 1, 2 or 3) identification of families in Eisen's designs	(Section 4)
R	number of replicates in Eisen's designs	(Section 4)
S	number of sires in Eisen's designs	(Section 4)
$\mathbf{S}_k$	sum of squares and products matrix	(Section 2)
s_k	size of $\mathbf{S}_k$	(Section 2)
s	number of matrices $\mathbf{S}_k$	(Section 2)
T_{ji}	coefficient of θ_i in Y_j	(Tables 1 and 7)
$\mathbf{U}$	matrices used to form $\mathbf{V}_k$	(A1)–(A6)
$\mathbf{V}_k$	expected value of $\mathbf{M}_k$	(Section 2)
x_{rqs}	measurement on sire in Eisen's design	(Appendix 2)
X_j	dummy for Y_j	(A9) and (A11)
Y_j	covariance between relatives	(Table 1)
y_{ijk}	measurement on P_i in family j from Bondari's design (k)	(Section 3)
y_{rqstd}	measurement on dam in Eisen's design	(Appendix 2)
z_{rqstdn}	measurement on offspring in Eisen's design	(Appendix 2)
θ_i	variance parameter	(Section 2)
σ^2_{XY}	variance component where X = A, D, E or P indicates additive, dominance, environmental or phenotypic component and Y = 0 or M indicates direct or maternal component, similarly the covariance component $\sigma_{X_0X_M}$	(Section 2)

Appendix 2

In this appendix we show how the results of section 2 can be applied to Eisen's designs (Section 4).

We let x_{rqs} represent the measurement on the sth sire from the rth replicate of family Q_q and y_{rqstd} represent the measurement on the dth dam from family Q_i mated to the sire with measurement x_{rqs}. t is a suffix used to identify the dam family. If the dam is in family Q_3 then t = 2. If the dam is in Q_1 or Q_2 t = 1. If t = 1 and q = 1 i.e., the sire is in Q_1 then the dam is in Q_2 and if q = 2 then the dam is in Q_1. We let z_{rqstdn} represent the nth offspring from the mating of parents with measurements x_{rqs} and y_{rqstd}. The ranges of the suffices are $r = 1 \cdots R$, $q = 1 \cdots S$, $d = 1 \cdots D_1$ (if t = 1), $d = 1 \cdots D_2$ (if t = 2) $n = 1 \cdots N$. Q = 1 for design I and Q = 2 for designs II and III. We let 0 indicate summation over all value of the particular suffix i.e. $z_{rqstdo} = \sum_{n=1}^{N} z_{rqstdn}$.

The data can be split into seven independent sum of squares and cross-product matrices. The sources of variation, the degrees of freedom (d_k) and size (s_k) of the matrices are as follows:

k	Source of variation	d_k	s_k
1	Within full-sibs	RQS $(D_1 + D_2)$ $(N - 1)$	1
2	Full-sibs (t = 1) within sires	RQS $(D_1 - 1)$	2
3	Full-sibs (t = 2) within sires	RQS $(D_2 - 1)$	2
4	Sires within sire families	RQ $(S - 1)$	5
5	Sire families within reps.	R $(Q - 1)$	5
6	Reps.	$R - 1$	5
7	(t = 1) and (t = 2)	1	2

to illustrate how to calculate the $\mathbf{S}_k$ we give the formulae for the elements of $\mathbf{S}_2$

$$S_{211} = \sum z^2_{rqsido} - (\sum z^2_{rqsioo})/D_1 ,$$

$$S_{221} = \sum z_{rqsido}y_{rqsid} - (\sum z_{rqsioo}y_{rqsio})/D_1 ,$$

$$S_{222} = \sum y^2_{rqsid} - (\sum y^2_{rqsio})/D_1 .$$

The other $\mathbf{S}_k$, like $\mathbf{S}_2$, are the differences between two matrices. The first matrix is the sum of squares and products of the vectors in Table 6. The second matrix is formed by first summing together l vectors

over a suffix m, and then forming the sum of squares and products of these summed vectors and dividing this matrix by l. The vectors suffices, m, and divisors l are indicated in Table 6. For instance for $\mathbf{S}_2$ the suffix is d and the divisor D_1.

The sum of squares matrix $\mathbf{S}_7$ is not quite the most efficient measure of differences between the two types of family but is suggested since it is easy to calculate and there is little information on the parameters from $\mathbf{S}_7$. We now give the expectation matrices $\mathbf{V}_k$ in terms of the covariances between relatives. The formulae for $\mathbf{V}_6$ and $\mathbf{V}_7$ are only correct if the families Q_i in the different replicates are unrelated. It is convenient to define several $\mathbf{U}$ matrices. Since the $\mathbf{U}$ matrices are symmetric only the lower triangles are given.

$$\mathbf{U}_1 = (Y_8) \tag{A1}$$

$$\mathbf{U}_2 = \begin{pmatrix} NY_8 + N^2(Y_7 - Y_3) & \\ N(Y_9 - Y_{11}) & Y_8 \end{pmatrix} \tag{A2}$$

$$\mathbf{U}_3 = \begin{pmatrix} NY_8 + N^2(Y_7 - X_4) & \\ N(Y_9 - X_{12}) & Y_8 + Y_7 - X_1 \end{pmatrix} \tag{A3}$$

$$\mathbf{U}_4 = \begin{pmatrix} U_{211} + N^2D_1^2(Y_3 - Y_5) & & & & \\ N^2D_1D_2(Y_1 - Y_2) & D_2U_{311} + N^2D_2^2(X_4 - X_6) & & & \\ D_1U_{221} & 0 & D_1U_{222} & & \\ 0 & D_2U_{321} & 0 & D_2U_{322} & \\ ND_1(Y_{10} - Y_{13}) & ND_2(Y_{10} - Y_{13}) & 0 & 0 & Y_8 \end{pmatrix} \tag{A4}$$

$$\mathbf{U}_5' = S\mathbf{U}_4 + S^2 \begin{pmatrix} N^2D_1^2Y_5 & & & & \\ N^2D_1D_2Y_2 & N^2D_2^2X_6 & & & \\ ND_1^2Y_{11} & 0 & D_1^2Y_7 & & \\ 0 & ND_2^2X_{12} & 0 & D_2^2X_1 & \\ ND_1Y_{13} & ND_2Y_{13} & 0 & 0 & Y_7 \end{pmatrix} \tag{A5}$$

Table 6

VECTORS, SUFFICES (l) AND DIVISORS (m) ASSOCIATED WITH $\mathbf{S}_k$ ($k = 1 \cdots 7$)

k	vector					ℓ	m
1	z_{rqstdk}					n	N
2	z_{rqsld0}	y_{rqsld}				d	D_1
3	z_{rqs2d0}	y_{rqs2d}				d	D_2
4	z_{rqsl00}	z_{rqs200}	y_{rqsl0}	y_{rqs20}	x_{rqs}	s	S
5	z_{rq0l00}	z_{rq0200}	y_{rq0l0}	y_{rq020}	x_{r0s}	q	Q
6	z_{r00l00}	z_{r00200}	y_{r00l0}	y_{r0020}	x_{r00}	r	R
7	z_{000t00}	y_{000t0}				t	2

$$\mathbf{U}_5'' = \begin{bmatrix} N^2D_1^2Y_{14} & & & & \\ N^2D_1D_2Y_{12} & N^2D_2^2X_{16} & & & \\ ND_1^2Y_{13} & ND_1D_2X_{13} & 0 & & \\ 0 & ND_2^2X_{12} & 0 & D_2^2X_1 & \\ ND_1Y_{11} & 0 & D_1Y_7 & 0 & 0 \end{bmatrix} \tag{A6}$$

$$U_{711} = R^2(U_{611}/D_1^2 - 2U_{621}/D_1D_2 + U_{622}/D_2^2)$$

$$U_{721} = R^2(U_{631}/D_1^2 - (U_{641} + U_{632})/D_1D_2 + U_{642}/D_2^2)$$

$$U_{722} = R^2(U_{633}/D_1^2 - 2U_{643}/D_1D_2 + U_{644}/D_2^2)$$

For design I

$$\mathbf{V}_1 = \mathbf{U}_1, \mathbf{V}_2 = \mathbf{U}_2, \mathbf{V}_3 = \mathbf{U}_3, \mathbf{V}_4 = \mathbf{U}_4, \mathbf{V}_6 = \mathbf{U}_6 = \mathbf{U}_5' \text{ and } \mathbf{V}_7 = \mathbf{U}_7 \tag{A8}$$

with

$$X_1 = Y_1, X_4 = Y_4, X_6 = Y_6, X_{12} = Y_{12} \text{ and } X_{16} = Y_{16}. \tag{A9}$$

For design II

$$\mathbf{V}_1 = \mathbf{U}_1, \mathbf{V}_2 = \mathbf{U}_2, \mathbf{V}_3 = \mathbf{U}_3, \mathbf{V}_4 = \mathbf{U}_4, V_5 = \mathbf{U}_5' - S^2\mathbf{U}_5''$$

$$\mathbf{V}_6 = \mathbf{U}_6 = 2\mathbf{U}_5' + 2S^2\mathbf{U}_5'' \text{ and } \mathbf{V}_7 = \mathbf{U}_7 \tag{A10}$$

with

$$X_1 = Y_7, X_4 = Y_3, X_6 = Y_5, X_{12} = Y_{11} \text{ and } X_{16} = Y_{15}. \tag{A11}$$

The $\mathbf{V}$'s of design III are defined as those of design II (A10) but the X's are defined as for design I (A9). In designs I and III the X's represent covariances with a paternal half-sib family, in design II the X's are associated with a full-sib family. Since $Y_j = \sum_{i=1}^{9} T_{ji}\theta_i$ we can find $\mathbf{F}_{ki}$ by evaluating $\mathbf{V}_k$ with Y_j replaced by T_{ji}.

Reprinted from
BIOMETRICS COPYRIGHT © 1976
THE BIOMETRIC SOCIETY, Vol. 32, No. 4, December 1976

15

Design of experiments to estimate heritability when observations are available on parents and offspring

by

R. Thompson

Reprinted from
BIOMETRICS COPYRIGHT © 1976
THE BIOMETRIC SOCIETY, Vol. 32, No. 2, June 1976

Design of Experiments to Estimate Heritability when observations are Available on Parents and Offspring

ROBIN THOMPSON

Statistical Laboratory, Iowa State University, Ames, Iowa
ARC Unit of Statistics, University of Edinburgh, 21 Buccleuch Place, Edinburgh EH8 9LN, Scotland[1]

Summary

The design of experiments to estimate heritability when data are available on both parents and offspring and the offspring data have a hierarchical structure is considered. Univariate maximum likelihood (ML) estimation is discussed, and extensions to the multivariate case are outlined. The efficiency of ML estimation is evaluated in cases where simple regression estimators are available. Optimum designs for ML estimation are given when various strategies of selecting and mating are followed. The variance of the heritability estimate can be approximately halved relative to designs in which no selection of parents is done.

1. Introduction

Hill and Nicholas [1974] have discussed the analysis and design of experiments to estimate heritability when data are available on both offspring and their parents. In particular, they compare estimation of heritability by regression and by sib-covariance with maximum likelihood (ML) estimation and show the advantages of using the ML procedure. The purpose of this paper is to consider the case when there is selection and assortative mating of the parents, which was not evaluated by Hill and Nicholas [1974]. Hill [1970] has enumerated the improvements in the sampling variances of regression estimates of heritability when there is selection and assortative mating of the parents.

In Section 2 we develop the relevant formulae in the univariate case for balanced designs when each sire is mated to several females. In Section 3 we compare the ML estimators with regression estimators for the situations discussed in Hill's 1970 paper. In Section 4 we consider optimal designs for estimating heritability and compare these with those of Hill and Nicholas [1974]. The emphasis is on making the most use of the data. It should be noted that the magnitude of the variances (or confidence intervals) of the resulting estimates depends on the total size of the experiment. Even if one of the designs advocated in this paper is used, there are circumstances when restrictions on the number of animals measured can mean that the confidence intervals are ridiculously large. In Section 5 we generalize these results to deal with p variates.

Journal paper No. J-8255 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa, Project No. 1669.

[1] Present address.

Key Words: Heritability estimation; Maximum likelihood estimation; Records subject to culling; Design of genetic experiments.

2. *Estimation from a Hierarchical Design*

We assume that we have observations on s' males and D' females and that the males and females are unrelated. Suppose s of the s' males each be mated to d dams with n offspring raised from each mating. Let x_i be the measurement on male $i(1 \leq i \leq s')$. Without loss of generality, we can let $1 \leq i \leq s$ represent males that have offspring. Let y_{ij} be the measurement on the jth dam mated to the ith sire $(1 \leq i \leq s, 1 \leq j \leq d)$, and z_{ijk} the measurement on her kth $(1 \leq k \leq n)$ offspring. Let $y_i(1 \leq i \leq D' - sd)$ represent the measurements on the females that have no recorded offspring (if $D' > sd$).

We assume that the observations are multivariate normally distributed with mean μ_1 for the males in the parent generation, mean μ_2 for the females in the parent generation, and mean μ_3 for the offspring. We assume that the variance of an observation is σ_P^2 irrespective of sex. We denote the additive genetic variance by σ_A^2. We assume that the covariance between parent and offspring is $\sigma_A^2/2$, again irrespective of sex. We assume that the covariance between full-sibs is $\sigma_A^2/2 + \sigma_K^2$. The covariance between full-sibs due to nonadditive genetic variation and to common environment is represented by σ_K^2. We assume that the covariance between half-sibs is $\sigma_A^2/4$. These assumptions are valid only if there are no interlocus interactions (for example epistasis) and the dam-offspring covariance is not valid if there is a maternal environmental covariance between parent and offspring.

This model for the covariance structure is equivalent to the one used by Hill and Nicholas [1974], but with a reparameterization. Their σ^2, H, K correspond to our $\sigma_P^2, \sigma_A^2/\sigma_P^2, \sigma_K^2/\sigma_P^2$, respectively. There are two reasons for the change in parameterization. Firstly, the linear parameterization $(\sigma_P^2, \sigma_A^2, \sigma_K^2)$ is simpler to work with and has more symmetry. Secondly, the linear parameterization case can easily be extended to deal with the multivariate case (Section 5).

In our parameterization heritability (h^2), the ratio of additive genetic variance to the total variance can be written as σ_A^2/σ_P^2, and similarly the proportion of variance due to common environment and dominance (K) can be denoted by σ_K^2/σ_P^2.

We first form various linear functions of the observations. These include $\bar{x}$, $\bar{y}$, $\bar{x}_1$, $\bar{y}_1$, $\bar{x}_2$, $\bar{y}_2$ and $\bar{z}_1$: $\bar{x}$ represents the mean of all male observations in the parent generation, $\bar{x}_1$ represents the mean of all males with offspring, and $\bar{x}_2$ represents the mean of all males with no offspring. $\bar{y}$, $\bar{y}_1$ and $\bar{y}_2$ are defined in a similar way for the female observations and $\bar{z}_1$ represents the mean of all offspring. Various family totals and means are required including y_{i0} and $y_{i.}$, the total and the mean value of the observations on females mated with the ith sire. Similarly z_{ij0} is the total for the offspring from the mating of the jth dam and ith sire, etc.

We let

$$
\begin{aligned}
s\bar{x}_1 &= \sum_{i=1}^{s} x_i , \qquad (s' - s)\bar{x}_2 = \sum_{i=s+1}^{s'} x_i , \qquad s'\bar{x} = \sum_{i=1}^{s'} x_i , \\
sd\bar{y}_1 &= \sum_{i=1}^{s} \sum_{j=1}^{d} y_{ij} , \qquad (D' - sd)\bar{y}_2 = \sum_{i=1}^{D'-sd} y_i , \\
D'\bar{y} &= sd\bar{y}_1 + (D' - sd)\bar{y}_2 , \qquad (1) \\
sdn\bar{z}_1 &= \sum_{i=1}^{s} \sum_{j=1}^{d} \sum_{k=1}^{n} z_{ijk}
\end{aligned}
$$

and

$$y_{i0} = \sum_{j=1}^{d} y_{ij}\,, \qquad dy_{i.} = y_{i0}\,,$$

$$z_{ij0} = \sum_{k=1}^{n} z_{ijk}\,, \qquad z_{i00} = \sum_{j=1}^{d} z_{ij0}\,,$$

$$nz_{ij.} = z_{ij0}\,, \qquad dz_{i.0} = z_{i00}\,.$$

We now form various quadratic functions of the data. For the data connected with parents and offspring (i.e., x_i , y_{ij} , z_{ijk} $i = 1, \cdots, s$, $j = 1, \cdots, d$, $k = 1, \cdots, n$), several functions are needed. For the offspring data, we define

$$\begin{aligned} M_{111} &= \sum_{i=1}^{s}\sum_{j=1}^{d}\sum_{k=1}^{n}(z_{ijk} - z_{ij.})^2 \\ M_{112} &= \sum_{i=1}^{s}\sum_{j=1}^{d}(z_{ij0} - z_{i.0})^2 \\ M_{113} &= \sum_{i=1}^{s}(z_{i00} - dn\bar{z}_1)^2, \end{aligned} \tag{2}$$

where M_{111} , M_{112} and M_{113} represents a subdivision of the offspring data into parts within full-sib families, between dams within sires, and between sires, respectively. For instance M_{111} , M_{112}/n and M_{113}/nd occur in a hierarchical analysis of variance of the offspring generation. Similarly, we let M_{222} and M_{223} represent subdivisions of the dam observations within and between sires respectively, and let M_{333} represent the sum of squares of males with offspring

$$\begin{aligned} M_{222} &= \sum_{i=1}^{s}\sum_{j=1}^{d}(y_{ij} - y_{i.})^2, \\ M_{223} &= \sum_{i=1}^{s}(y_{i0} - d\bar{y}_1)^2, \\ M_{333} &= \sum_{i=1}^{s}(x_i - \bar{x})^2. \end{aligned} \tag{3}$$

In a similar way, we can define cross-products between the offspring and females both within sires (M_{122}) and between sires (M_{123}), and also between sires with offspring (M_{133}) and with females (M_{233})

$$\begin{aligned} M_{122} &= \sum_{i=1}^{s}\sum_{j=1}^{d}(z_{ij0} - z_{i.0})(y_{ij} - y_{i.}), \\ M_{123} &= \sum_{i=1}^{s}(z_{i00} - dn\bar{z})(y_{i0} - d\bar{y}_1), \\ M_{133} &= \sum_{i=1}^{s}(z_{i00} - dn\bar{z})(x_i - \bar{x}_1), \\ M_{233} &= \sum_{i=1}^{s}(y_{i0} - d\bar{y}_1)(x_i - \bar{x}_1). \end{aligned} \tag{4}$$

The suffixes on M_{ijk} ($k = 1, 2, 3$) are chosen so that i and j indicate which part of the data is involved in the cross-products, i.e., offspring (1), female parents (2), or male parents

(3), and k indicates the subdivision into within full-sib families (1), between dams within sires (2), and between sires (3).

Another sum of squares is needed, M_{114}, and is made up of four parts. One part is the sum of squares of males with no offspring, another is the sum of squares for the difference between $\bar{x}_1$ and $\bar{x}_2$, i.e., between males with offspring and without offspring. There are two corresponding terms for the females

$$M_{114} = \sum_{i=s+1}^{s'} (x_i - \bar{x}_2)^2 + \frac{s(s'-s)}{s'}(\bar{x}_1 - \bar{x}_2)^2 + \sum_{i=1}^{D} (y_i - \bar{y}_2)^2 + \frac{sd(D'-sd)}{D'}(\bar{y}_1 - \bar{y}_2)^2. \quad (5)$$

We can form four symmetric matrices $\mathbf{M}_k$ ($k = 1, \cdots, 4$) of size $r_k \times r_k$ (1, 2, 3, 1) with elements i, j of $\mathbf{M}_k$ equal to M_{ijk}. If $s' = s$ and $D' = sd$, then $\mathbf{M}_4$ is null, and, similarly, if $n = 1$, $d = 1$, or $s = 1$, the matrices $\mathbf{M}_1$, $\mathbf{M}_2$ or $\mathbf{M}_3$ are null. The expected value of M_k can be written as $\nu_k V_k$, where ν_k is the degrees of freedom associated with $\mathbf{M}_k$, i.e., $\nu_1 = sd(n-1)$, $\nu_2 = s(d-1)$, $\nu_3 = (s-1)$, and $\nu_4 = s' - s + D' - sd$ and

$$\mathbf{V}_k = \sum_{i=1}^{3} \theta_i \cdot f_k(\theta_i)$$

where $\theta_1 = \sigma_P^2$, $\theta_2 = \sigma_A^2$, and $\theta_3 = \sigma_K^2$. The f matrices can be found to be

$$f_1(\sigma_P^2) = (1), \quad f_1(\sigma_A^2) = (-1/2), \quad f_1(\sigma_K^2) = (-1)$$

$$f_2(\sigma_P^2) = \begin{bmatrix} n & 0 \\ 0 & 1 \end{bmatrix}, \quad f_2(\sigma_A^2) = \begin{bmatrix} n(n-2)/4 & n/2 \\ n/2 & 0 \end{bmatrix}, \quad f_2(\sigma_K^2) = \begin{bmatrix} n(n-1) & 0 \\ 0 & 0 \end{bmatrix}$$

$$f_3(\sigma_P^2) = \begin{bmatrix} nd & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad f_3(\sigma_A^2) = \begin{bmatrix} nd(nd+n-2)/4 & nd/2 & nd/2 \\ nd/2 & 0 & 0 \\ nd/2 & 0 & 0 \end{bmatrix},$$

$$f_3(\sigma_K^2) = \begin{bmatrix} nd(n-1) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$f_4(\sigma_P^2) = (1)$, $f_4(\sigma_A^2) = (0)$, and $f_4(\sigma_K^2) = (0)$.

Hill and Nicholas [1974] use terms similar to $\mathbf{V}_1$, $\mathbf{V}_2$ and $\mathbf{V}_3$. If we reorder the rows and the columns of their $\mathbf{S}_2$ to correspond to the order used in $\mathbf{V}_2$ we find

$$\begin{bmatrix} n & 0 \\ 0 & 1 \end{bmatrix} \mathbf{S}_2 \begin{bmatrix} n & 0 \\ 0 & 1 \end{bmatrix} \sigma_P^2 = \mathbf{V}_2,$$

if we take account of the different parameterization of Hill and Nicholas. There is a similar relationship between $\mathbf{V}_3$ and $\mathbf{S}_1$. The $\mathbf{M}_k$ have been defined in this paper in order to avoid the fractions of n and nd that occur in Hill and Nicholas' paper.

The f matrices enable the expected values of M_{ijk} to be written in a compact linear form. They are analogous to the elements of the design matrix $\mathbf{X}$ in the usual linear model formulation $E(\mathbf{y}) = \mathbf{X}\beta$. They indicate for instance that

$$E(M_{111}) = sd(n-1)[\sigma_P^2 - \sigma_A^2/2 - \sigma_K^2]$$

and

$$E(M_{113}) = (s - 1)[nd\sigma_P^2 + nd(nd + n - 2)\sigma_A^2/4 + nd(n - 1)\sigma_K^2].$$

Suppose first that parents are selected at random. To estimate σ_P^2, σ_A^2 and σ_K^2, following Patterson and Thompson [1971], we maximize not the total likelihood but that of error contrasts; i.e., contrasts that tell us nothing about the fixed effects μ_1, μ_2 and μ_3.

By partitioning the likelihood in a similar way to that used in defining $\mathbf{M}_k$ (i.e., into parts connected with the parent and offspring data) with the rest of the data and with contrasts between the two parts, the log likelihood corresponding to the error contrasts can be shown to be

$$\text{const} - (1/2) \sum_{k=1}^{4} (\nu_k \ln |\mathbf{V}_k| + \text{tr}\,(\mathbf{M}_k\mathbf{V}_k^{-1})), \tag{6}$$

where $|\mathbf{A}|$ and tr $(\mathbf{A})$ represent the determinant and trace of the square matrix $\mathbf{A}$.

The information matrix of θ_1, θ_2 and θ_3 is the matrix $\mathbf{G}$ with elements

$$G_{ij} = (1/2) \sum_{k=1}^{4} \nu_k \text{ tr } [(\mathbf{V}_k^{-1}f_k(\theta_i))(\mathbf{V}_k^{-1}f_k(\theta_j))]. \tag{7}$$

Once estimates of σ_P^2, σ_A^2 and σ_K^2 have been found, we can estimate μ_1, μ_2 and μ_3 as

$$\hat{\mu}_1 = \bar{x}$$

$$\hat{\mu}_2 = \bar{y}$$

$$\hat{\mu}_3 = \bar{z} - (\bar{y}_1 - \bar{y} + \bar{x}_1 - \bar{x})(\hat{\sigma}_P^2)^{-1}\hat{\sigma}_A^2.$$

If parents are selected (or mated assortively) on the basis of their records, we find, by considering the likelihood of the parent generation's records and the likelihood of the offspring conditional on their parent's records, that the log likelihood can still be written as (6), at least for the simple model of inheritance assumed in this paper. A similar argument is used in Henderson *et al.* [1959].

However, the expectation of the moments M_{222}, M_{223}, M_{333} and M_{233}, or equivalently the sums of squares of dams within and between sires, the sum of squares between sires, and the sum of products between sires and their mates, which are used in calculating the information matrix, do depend on the selection scheme and we write the information matrix conditional on these values. (Curnow [1961] has used a similar argument.) It is useful to define terms M_1^*, M_2^*, M_3^*, M_{23}^* and M_{32}^*, which arise in the modified information matrix. Thompson [1973] used similar terms in a multivariate context. We let

$$M_1^* = (\sigma_P^2)^{-1}(M_{222} - \nu_2\sigma_P^2)(\sigma_P^2)^{-1},$$

$$M_2^* = (\sigma_P^2)^{-1}(M_{223} - \nu_3 d\sigma_P^2)(\sigma_P^2)^{-1},$$

$$M_3^* = (\sigma_P^2)^{-1}(M_{333} - \nu_3\sigma_P^2)(\sigma_P^2)^{-1}.$$

These are measures of the selection carried out on dams within sires (M_1^*), on dams between sires (M_2^*), and on sires (M_3^*), respectively. For instance, if the highest and lowest yielding males are used as sires, the sires sum of squares, M_{333}, will be greater than $\nu_3\sigma_P^2$ and M_3^* will be positive. On the other hand, if sires are selected at random from the males in the parent generation, then M_{333} has expected value $\nu_3\sigma_P^2$ and so the expected value of M_3^* is zero. In fact, if parents with recorded offspring are chosen at random, then the expected values of M_1^*, M_2^* and M_3^* are zero. Similarly, we let

$$M_{23}^* = (\sigma_P^2)^{-1} M_{233} (\sigma_P^2)^{-1}$$

and

$$M_{32}^* = (\sigma_P^2)^{-1} M_{323} (\sigma_P^2)^{-1}.$$

These are measures of the assortative mating carried out; for instance, if mates are chosen at random the expected values of M_{23}^* and M_{32}^* are zero.

The information matrix $\mathbf{G}^*$ is modified and has elements

$$G_{ij} + G_{ij1} + G_{ij2}$$

where

$$\begin{aligned} G_{ij1} &= \tfrac{1}{4}(-1)^{i+j}(n^2 M_1^*) \cdot \theta_i^{-1} \sigma_A^2 V_2^{*-1} \sigma_A^2 \theta_j^{-1}, \\ G_{ij2} &= \tfrac{1}{4}(-1)^{i+j}(n^2 M_2^* + n^2 d M_{23}^* + n^2 d^2 M_{32}^* + n^2 d^2 M_3^*) \cdot \theta_i^{-1} \sigma_A^2 V_3^{*-1} \sigma_A^2 \theta_j^{-1} \end{aligned} \tag{8}$$

for $i, j = 1$ and 2 with G_{ij1} and $G_{ij2} = 0$ if i or $j = 3$

and

$$\begin{aligned} V_2^* &= \sum_{i=1}^{3} \theta_i \cdot f_{112}(\theta_i) - \left(\frac{n^2}{4}\right) \sigma_A^2 (\sigma_P^2)^{-1} \sigma_A^2, \\ V_3^* &= \sum_{i=1}^{3} \theta_i \cdot f_{113}(\theta_i) - \left(\frac{n^2}{4} + \frac{n^2 d}{4}\right) \sigma_A^2 (\sigma_P^2)^{-1} \sigma_A^2, \end{aligned} \tag{9}$$

$f_{112}(\theta_i)$ and $f_{113}(\theta_i)$ are the elements in the first row and column of $f_2(\theta_i)$ and $f_3(\theta_i)$, respectively.

The derivations of these results are straightforward but tedious (some details are in the appendix). The $(-1)^i (\sigma_P^2)^{-1} \theta_i^{-1} \sigma_A^2$ terms arise from differentiating the regression of offspring records on their parents, which include terms in $(\sigma_P^2)^{-1} \sigma_A^2$. V_2^* and V_3^* can be interpreted quite simply in terms of the variance of the offspring's records given the parent's records. The results can be modified in a fairly obvious manner if observations are not available on the female or male parents. For instance, if females are not measured, $r_2 = 1$, $r_3 = 2$; we can delete the relevant rows and columns from $\mathbf{M}_k$ and $\mathbf{d}_k$, $G_{ij1} = 0$, and we lose the $(n^2 d \sigma_A^2 (\sigma_P^2)^{-1} \sigma_A^2/4)$ term from V_3^*. Similarly, if sires are not measured, $r_3 = 2$, and we can delete the relevant rows and columns from the matrices and we lose the

$$(n^2 \sigma_A^2 (\sigma_P^2)^{-1} \sigma_A^2/4)$$

term from V_3^*. If both dam and sire observations are not available, then only three elements of the $\mathbf{M}$'s exist and correspond to terms in the hierarchical analysis of variance of the offspring observations.

3. Comparison with Regression Estimators

In this section, we compare the efficiency of the ML estimator of heritability regression estimators for the situations discussed by Hill [1970]. He considers four situations where simple regression estimators of heritability are available:

1. When equal numbers of males and females are measured and each male is mated to only one female.
2. When only one parent is measured.
   a. Offspring families are half-sibs and related only through the measured parent.
   b. Each measured parent has only one mate so that the offspring are full-sibs.
   c. All measured parents have the same mate.

In Case 1, Hill considered four selection and mating strategies for parents.

(i) Random mating.
(ii) Assortative mating.
(iii) Selection of the highest and lowest p percent of males and females and random mating within high and low groups.
(iv) Selection of the highest and lowest p percent of males and females and assortative mating.

In Case 2, two strategies were considered.

(i) Using all measured parents.
(ii) Selection of the highest and lowest p percent of the parents.

In the notation of this paper we have

in Case 1,

$$r_2 = 2, \quad r_3 = 3, \quad s = ?, \quad d = 1, \quad n = ?, \quad s' = s/2p, \quad D' = s/2p,$$

in Case 2a,

$$r_2 = 1, \quad r_3 = 2, \quad s = ?, \quad d = ?, \quad n = 1, \quad s' = s/2p,$$

in Case 2b,

$$r_2 = 2, \quad r_3 = 2, \quad s = ?, \quad d = 1, \quad n = ?, \quad s' = s, \quad D' = sd/2p,$$

in Case 2c,

$$r_2 = 2, \quad r_3 = 2, \quad s = 1, \quad d = ?, \quad n = ?, \quad s' = s, \quad D' = sd/2p.$$

When $p = 0.5$ and there is no selection of parents, ? indicates a parameter not specified by the mating strategy. For instance, when both parents are measured, we can form all the elements of $\mathbf{M}_2$ and $\mathbf{M}_3$, so that $r_2 = 2$ and $r_3 = 3$. Each sire is mated to one dam so that $d = 1$. The total number of females, D', is equal to s', the total number of males. The number of males used as parents, s, is equal to $2ps'$, so that a proportion $2p$ of the males are used as parents. Similarly, $2ps'$ females are used as dams.

The values of $\mathbf{M}_2$ and $\mathbf{M}_3$ used in the evaluation of (8) are when both parents are measured, (Case 1),

(i) $M_{223} = M_{333} = (s - 1)\sigma_P^2, \quad M_{233} = M_{323} = 0,$

(ii) $M_{223} = M_{333} = (s - 1)\sigma_P^2, \quad M_{233} = M_{323} = (s - 1)\sigma_P^2,$

(iii) $M_{223} = M_{333} = (s - 1)(1 + ix)\sigma_P^2, \quad M_{233} = M_{323} = (s - 1)i^2\sigma_P^2,$

(iv) $M_{223} = M_{333} = (s - 1)(1 + ix)\sigma_P^2, \quad M_{233} = M_{323} = (s - 1)(1 + ix)\sigma_P^2,$

where x and i are the abscissa and the selection differential on the standardized normal curve when the top p percent are chosen.

We note that the use of x and i in this section should not be confused with their use in the previous section, where x_i denoted an observation on a male and i was used as an index. Similar formulae hold when only one parent is measured. When both parents are measured, the sums of squares for midparents $(M_{223} + M_{233} + M_{323} + M_{333})/4$ become for the four strategies

(i) $(s - 1)(\sigma_P^2/2),$

(ii) $(s - 1)\sigma_P^2,$

(iii) $(s - 1)(1 + ix + i^2)(\sigma_P^2/2),$

(iv) $(s - 1)(1 + ix)\sigma_P^2.$

Table 1

OPTIMUM FAMILY SIZE m, SAMPLING VARIANCE $v(\times 10^2)$ AND OPTIMUM PROPORTION ($p\%$) FOR ALTERNATIVE METHODS FOR VARIOUS VALUES OF h^2.

h^2	0.05			0.10			0.20			0.40			0.60			0.80		
Case	m	v	p	m	v	p	m	v	p	m	v	p	m	v	p	m	v	p
1(i)	39	77		19	144		9	247		4	330		3	291		4	181	
1(ii)	33	64		15	105		7	150		4	178		3	160		4	110	
1(iii)	38	44	6.6	21	68	8.0	12	96	9.6	8	119	11.1	7	110	11.8	8	76	11.5
1(iv)	37	44	6.9	20	66	8.5	12	94	10.2	8	115	11.7	7	107	12.4	8	74	12.1
2a(i)	70	151		30	277		10	445		4	551		4	524		7	395	
2a(ii)	65	99	3.2	32	150	4.1	17	212	5.4	11	262	6.4	10	253	6.7	14	190	5.8
2b(i)	40	77		20	146		10	261		5	413		3	476		3	469	
2b(ii)	41	69	8.5	21	123	9.2	12	201	9.7	7	293	10.7	5	328	11.9	4	324	12.9
2c(i)	70	145		29	259		10	398		4	451		3	386		3	271	
2c(ii)	64	96	3.2	32	142	4.1	17	194	5.3	10	225	6.7	9	204	6.9	11	143	6.3

These are of the same form as the formulae derived by Hill [1970] (except that he has a factor s instead of $(s - 1)$).

In order to make a fair comparison with Hill's results, we first assume, as he did, that $\sigma_K^2 = 0$; i.e., that the covariance between full-sibs arises solely from additive genetic variance. We will relax this assumption later. In Table 1 are given the balanced designs that minimize v, the variance of the estimate of heritability per observation, that is, $v = V(h^2)$. T, where $V(h^2)$ is the variance of the estimate of heritability and T is the total number of observations in the experiment. $V(h^2)$ was calculated by using the usual Taylor series approximation for the variance of a ratio. The variances and covariances of σ_A^2 and σ_P^2 were found from the inverse of the information matrix **G**. T can be written as $sdn + s'(r_3 - r_2) + D'(r_2 - 1)$. The optimal designs are essentially independent of s (or d in Case 2c), and for convenience v has been calculated with $s = 100$ ($d = 100$ for Case 2c). The family size, m, is equivalent to n in Cases 1, 2b and 2c and to d in Case 2a.

At moderate and high values of h^2, the family sizes are similar to those suggested by Hill, and the values of v are of the order of 10 percent to 30 percent smaller than those given by Hill, there being roughly half as much reduction in v when the parents are selected compared with when they are not. There is a larger reduction in v at low values of heritability, and family sizes are much larger than those given by Hill. This might be expected since at low values of heritability the sib correlation information is more valuable than the regression information, and Robertson [1959] has shown that relatively large family sizes are needed to estimate heritability efficiently with sib correlation methods. The case when only one parent is measured and the progeny are half-sibs and there is no selection of parents (2$a(i)$) was considered by Hill and Nicholas [1974]. At low values of heritabilities, Cases 2a and 2c are similar because most of the information comes from the sib-covariance part and this in both cases estimates $\sigma_A^2/4$.

In Cases 1 and 2b, the offspring are full-sibs and the covariance structure of families is similar to the covariance structure of $n + 2$ and $n + 1$ full-sibs. The only difference is that when both parents are measured the covariance between the sire and dam is zero and not

$\sigma_A^2/2$. Robertson [1959] has shown that for data on full-sibs alone the optimum family size is approximately $2/h^2$. A better approximation is $1 + 2/h^2$, so the optimum family size becomes approximately $2/h^2 - 1$ and $2/h^2$ in the two cases. These values are in good agreement with the values in Table 1 when there is no selection or assortative mating of parents. The optimum proportions p are similar to the values found by Hill.

Hill [1970] only considers the case when σ_K^2 is assumed to be zero. When σ_K^2 is non-zero, the variance of the regression estimate of heritability per observation, v_R , becomes

$$v_R = \frac{A(Bnp + 1)(1 - h^2/2 - K + nC)}{npD} \tag{10}$$

where $A = 1$ in Cases 1(i) and 1 (iii),
$A = 2$ in other cases,
$B = 1$ when both parents are measured,
$B = 2$ when one parent is measured,
$C = (h^2/2 - h^4/2 + K)$ in Case 1,
$C = (h^2/2 - h^4/4 + K)$ in Case 2*b*,
$C = (h^2/4 - h^4/4 + K)$ in Case 2*c*,
$D = (1 + ix + i^2)$ in Case 1 (iii),
$D = (1 + ix)$ in other cases.

Case 2*a* will not be considered in this sub section since, in this situation, v_R is then independent of σ_A^2 and σ_K^2 cannot be estimated. Formulae (10) reduces to Hill's formulae when $\sigma_K^2 = 0$ and to those Hill and Nicholas [1974] when there is no selection.

Differentiating v_R with respect to n and p and equating the differentials to zero we get

$$n^2p = \frac{1 - h^2/2 - K}{BC} \tag{11}$$

and

$$n = \frac{x^2}{p(1 + ix - x^2)} ,$$

in Cases 1 (i), 1 (ii) and 1 (iv),

$$n = \frac{x^2 + 2ix - i^2}{p(1 - x^2 - ix + 2i^2)} , \tag{12}$$

in Case 1 (iii)
and

$$n = \frac{x^2}{2p(1 + ix - x^2)}$$

in Case 2.

When $K = 0$, equations (11) and (12) reduce to some given by Hill [1970]. Solutions to (12) for p in terms of n have been tabulated by Hill.

In these situations there is little gain by using the ML estimator compared with the regression estimator. When σ_K^2 was assumed to be zero, the advantage from using the ML estimator arose essentially as a result of the use of the sib-covariance data; however, in this subsection this full-sib covariance data is used to estimate σ_K^2.

In Table 2 are given the optimal family sizes for balanced designs by using ML estimation in six cases (1(i), 1(iv), 2*b*(i), 2*b*(ii), 2*c*(i) and 2*c*(ii)) and for $K = 0.05$ and $K = 0.2$.

Table 2

OPTIMUM FAMILY SIZE (n) WHEN K IS ESTIMATED FOR CASES 1(i), 1(iv), 2b(i), 2b(ii), 2c(i) AND 2c(ii) FOR VARIOUS VALUES OF h^2 AND K.

		h^2					
Case	K	0.05	0.10	0.20	0.40	0.60	0.80
1(i)	0.05	5	4	4	3	3	3
1(iv)		12	10	7	6	5	6
2b(i)		4	3	2	2	2	2
2b(ii)		9	8	6	4	3	3
2c(i)		4	3	3	2	2	2
2c(ii)		11	9	8	6	5	5
1(i)	0.20	3	3	2	2	2	2
1(iv)		5	5	4	3	3	3
2b(i)		2	2	2	2	2	2
2b(ii)		4	4	3	2	2	2
2c(i)		2	2	2	2	2	2
2c(ii)		4	4	3	3	3	2

A good approximation for p is given by (11). The values of n tabulated are close to the solutions of (10) and (11). Optimal designs for $K = 0$ correspond to the designs suggested by Hill. As K increases, the optimal family size decreases. The reduction in v by selection and assortative mating of parents also decreases as K increases, but is still appreciable. For instance, in Case 2*b* the reduction is of the order of a half when $h^2 = 0.2$ and $K = 0$, and of the order of a third when $h^2 = 0.2$ and $K = 0.2$.

Family sizes of 3, 6, 3, 4, 3 and 6 for the Cases 1(i), 1(iv), 2*b*(i), 2*b*(ii), 2*c*(i) and 2*c*(ii) (with selection of parents, $p = 10$ percent) were found to be reasonably efficient over a range of values of h^2 and K. These designs could be useful if there was very vague knowledge on the probable values of h^2 and K before an experiment started.

4. Other Hierarchical Designs

In the previous section, designs that gave rise to simple regression estimators of heritability were considered. In this section, four other cases are considered. Case 3, measurements on males and females,

(i) The highest and lowest p percent of males are selected and used as sires,

(ii) An equal number of males and females are measured, the highest and lowest p percent of males and dp percent of females are selected. Mating is at random with the high and low selected groups.

Case 4(5), measurements on males (females) alone, selection of the highest and lowest p percent of males (females) and mating at random.

In the notation of the paper we have

in Case 3(i), $r_2 = 2$, $r_3 = 3$, $s' = s/2p$, $D' = sd$,

in Case 3(ii), $r_2 = 2$, $r_3 = 3$, $s' = s/2p$, $D' = s/2p$,

in Case 4, $r_2 = 1, \quad r_3 = 2, \quad s' = s/2p,$

in Case 5, $r_2 = 2, \quad r_3 = 2, \quad s' = s, \quad D' = sd/2p.$

For example, when only females are measured, the terms M_{133}, M_{233} and M_{333} related to the sire part of the data do not exist, so that $r_3 = 2$. All the terms of M_2 are defined so $r_2 = 2$. There are no measured males that do not have offspring, so $s' = s$, and there is a total of D' females and sd(a proportion $2p$) that have offspring. The terms in (8) were evaluated by using, when both parents are measured and males are selected (Case 3(i)),

$$M_{222} = s(d-1)\sigma_P^2, \qquad M_{233} = M_{323} = 0,$$
$$M_{223} = d(s-1)\sigma_P^2, \quad \text{and} \quad M_{333} = (s-1)(1+ix)\sigma_P^2,$$

and when both parents are measured and selected (Case 3(ii)),

$$M_{222} = s(d-1)(1 - i_d(i_d - x_d))\sigma_P^2, \qquad M_{233} = M_{323} = (s-1)di_d i\sigma_P^2,$$
$$M_{223} = d((sd-1)i_d^2 + (s-1)(1 - i_d(i_d - x_d)))\sigma_P^2$$

and

$$M_{333} = (s-1)(1+ix)\sigma_P^2,$$

where i_d and x_d are the abscissa and selection differential when the top pd percent are selected from the standardized normal curve.

For instance, in Case 3(ii) the variance of dams within selected groups (high and low) is $(1 - i_d(i_d - x_d))\sigma_P^2$ (for example, Hill [1970]), so that M_{222} is $s(d-1)(1 - i_d(i_d - x_d))\sigma_P^2$ and the total sum of squares for dams $(= M_{222} + M_{223}/d)$ is $(sd-1)(1 + i_d x_d)\sigma_P^2$. When males alone are measured (Case (4))

$$M_{222} = M_{223} = M_{233} = M_{323} = 0,$$
$$M_{333} = (s-1)(1+ix)\sigma_P^2.$$

and when females alone are measured (Case (5))

$$M_{222} = s(d-1)(1+ix)\sigma_P^2, \qquad M_{233} = ds(1+ix)\sigma_P^2$$

and

$$M_{233} = M_{323} = M_{333} = 0.$$

The balanced designs that minimize v have been found for these cases and are given in Table 3 when both parents are measured and when females alone are measured (Cases 3 and 5). The situation when males alone are measured (Case 4) is very simple and does not need tabulating. In this case, the optimal designs correspond to $n = 1$ and hence are covered by Case 2a in Section 3. If, however, σ_K^2 needs to be estimated, $n = 2$ is optimal and d is approximately half the value for Case 2a(ii); that is, the sire family size, dn, stays approximately constant. The optimum proportion selected is of the same order as given for Case 2a(ii).

As noted by Hill and Nicholas, who considered the case of no selection or assortative mating, it is seen that for constant h^2 the optimum value of dn is effectively independent of K when both parents are measured. When there is selection of males, the values of d and n are similar to those given by Hill and Nicholas. However, selection of males reduces v

Table 3

OPTIMUM FAMILY STRUCTURE (d, n) FOR CASES 3(i), 3(ii), AND 5, FOR VARIOUS VALUES OF h^2 AND K.

		h^2					
Case	K	0.05	0.10	0.20	0.40	0.60	0.80
3(i)		10, 8	7, 6	5, 4	5, 3	4, 3	5, 3
3(ii)	0.0	2,21	2,13	1,11	1, 8	1, 7	1, 8
5		7,20	7,15	10,11	*,10	*, 9	*,11
3(i)		19, 4	12, 3	7, 3	7, 2	6, 2	7, 2
3(ii)	0.05	7, 7	4, 6	2, 6	2, 4	2, 4	2, 4
5		19, 6	11, 7	10, 6	*, 6	*, 5	*, 5
3(i)		39, 2	20, 2	10, 2	7, 2	7, 2	7, 2
3(ii)	0.20	17, 3	8, 3	6, 2	4, 2	4, 2	4, 2
5		50, 2	22, 3	15, 3	17, 3	32, 2	*, 2

* indicates that the optimum value of d is greater than 99.

by about a third relative to the designs where no selection is practiced. The optimum proportions of males selected are again close to the values tabulated for Case 2*a*(ii) in Table 1.

When there is selection of males and females (Case 3(ii)), the values of dn are approximately half of their corresponding value when only males are selected, and v is reduced by about half compared with the designs given by Hill and Nicholas. When $K = 0$, the optimal proportion selected is approximately given by

$$dn = \frac{x^2 + 2ix - i^2}{p(1 - x^2 - ix + 2i^2)},$$

which corresponds to (12) (Case 1(iii)) with dn replacing n. Optimal selection proportions decrease as K increases, being, for example, 3.7 percent when $h^2 = 0.1$ and $K = 0.2$ and 7.6 percent when $h^2 = 0.6$ and $K = 0.2$.

When only females are measured, the situation is more complicated. The tabulated values of n are close to the values of n when both parents are measured and selected, but the values of d are larger. The larger values of d in the table might be explained intuitively as follows: At moderate and large values of h^2 most of the information on h^2 comes from the regression type estimators, particularly the within-sire regression estimate, and following the arguments of Latter and Robertson [1960], for example, it can be seen that the variance per observation of the intra-sire regression estimate of heritability is a monotone decreasing function of d. Further evidence for this is to note that, when females are measured and mated to one sire (Case 2*c*), the only regression type estimator available is the within sire dam-daughter regression and that, when the optimal d in Case 5 is large, the optimal n coincides with the corresponding value in Case 2*c*(ii). The large values of d are obviously not practical, but reducing d from 95 to 35 only increases v by about one percent, for those designs that have an optimal d greater than 99. At moderate to high values of h^2, the optimal proportion of females selected is of the order of 7, 9 and 15 percent for $K = 0.0$, 0.05 and 0.20, respectively, and when $h^2 = 0.05$, the respective values are 7, 19 and 50 percent.

Table 4

OPTIMUM SIZE (d) AND PROPORTION (p%) FOR CASES 3(i), 3(ii) AND 5, WHEN OBSERVATIONS ARE AVAILABLE ON PARENTS AND HALF-SIBS ONLY, FOR VARIOUS VALUES OF h^2.

	h^2											
Case	0.05		0.10		0.20		0.40		0.60		0.80	
3(i)	61	2.2	30	3.2	17	4.5	11	6.0	10	6.6	11	6.6
3(ii)	38	1.3	17	2.4	10	4.7	6	6.9	6	7.1	7	6.7
5	82	50.0	43	50.0	24	34.4	18	29.0	21	28.3	37	29.0

Reductions in v when compared with designs where there is no selection vary from 0 percent (when $h^2 = 0.05$, $K = 0.20$) to 50 percent (when $h^2 = 0.6$, $K = 0.0$).

Sometimes only half-sib families are possible in the offspring generation, that is $n = 1$, the optimum family size (d) and proportion (p percent) are given in Table 4, for Cases 3 and 5. When both parents are measured, the reduction in v compared with designs in which no selection is practiced are similar to those when full-sib families were possible. When only females are measured, there is very little gain in selecting parents; in fact, at low values of h^2 the optimum procedure is to use all the females as parents.

When little is known about h^2 and K, reasonable compromise designs seem to be

$$d = 10, \quad n = 2 \quad p = 5\% \quad \text{in Case 3(i)}$$

$$d = 5, \quad n = 4 \quad p = 15\% \quad \text{in Case 3(ii)}$$

$$d = 10, \quad n = 6 \quad p = 10\% \quad \text{in Case 5,}$$

and when only half-sib families are possible

$$d = 24 \quad p = 5\% \quad \text{in Case 3(i)}$$

$$d = 10 \quad p = 15\% \quad \text{in Case 3(ii)}$$

and

$$d = 24 \quad p = 50\% \quad \text{in Case 5.}$$

5. *Multivariate Generalization*

If we have observations on p variates then it can be of some interest to estimate the variance and covariance components. The preceding results can be generalized to cover this and can be expressed compactly by using matrix operators defined and discussed by Thompson [1973]. Instead of estimating scalars σ_P^2, σ_A^2 and σ_K^2 representing variances, we estimate symmetric $p \times p$ matrices $\boldsymbol{\sigma}_P$, $\boldsymbol{\sigma}_A$ and $\boldsymbol{\sigma}_K$ representing variances and covariances. Vectors of means can be defined analogous to (1). We can form matrices $\mathbf{M}_k^{rs}$ representing crossproducts between the rth and sth variates by using formulae (2) – (5). Symmetric matrices $\mathbf{M}_k$ of size (ps_k) can be formed with (r, s)th submatrix $\mathbf{M}_k^{rs}$, $\mathbf{V}_k$ becomes

$$\mathbf{V}_k = \sum_{i=1}^{3} \boldsymbol{\theta}_i * f_k(\boldsymbol{\theta}_i)$$

where * represents the left direct product; see, for instance, Searle [1966].

To illustrate the formation of these matrices, we adopt a numerical example taken from Thompson [1973]. He gives data that can be thought of arising when 1,000 females are measured in the parent generation, one offspring is raised from 501 of these females and two variates are measured so that in the notation of this paper,

$$p = 2,\ D' = 1000,\ d = 1,\ n = 1,\ s' = s = 501.$$

Because only females are measured in the parent generation, $r_2 = 2$ and $r_3 = 2$. The degrees of freedom associated with $\mathbf{M}_k$ $(k = 1, \cdots, 4)$ are

$$\nu_1 = 0,\ \nu_2 = 0,\ \nu_3 = 500,\ \nu_4 = 499.$$

M_1 and M_2 are null because only one offspring is raised from each mating and each dam is assumed to have mated to a different sire. The M_3 matrix for the first variate (taken from Thompson [1973]) is

$$\mathbf{M}_3^{11} = \begin{bmatrix} 466.72 & 36.85 \\ 36.85 & 175.12 \end{bmatrix},$$

representing the sum of squares for offspring (466.72), the sum of squares for dams (175.12) and the sum of crossproducts for offspring with dams (36.85) for the first variate. Similarly we find,

$$\mathbf{M}_3^{12} = \begin{bmatrix} 207.43 & 3.18 \\ 30.06 & 92.57 \end{bmatrix},$$

$$\mathbf{M}_3^{21} = \begin{bmatrix} 207.43 & 30.06 \\ 3.18 & 92.57 \end{bmatrix},$$

$$\mathbf{M}_3^{22} = \begin{bmatrix} 471.82 & 95.70 \\ 95.70 & 446.72 \end{bmatrix}.$$

The composite $\mathbf{M}_3$ becomes

$$\mathbf{M}_3 = \begin{bmatrix} \mathbf{M}_3^{11} & \mathbf{M}_3^{12} \\ \mathbf{M}_3^{21} & \mathbf{M}_3^{22} \end{bmatrix} = \begin{bmatrix} 466.72 & 36.85 & 207.43 & 3.18 \\ 36.85 & 175.12 & 30.06 & 92.57 \\ 207.43 & 30.06 & 471.82 & 95.70 \\ 3.18 & 92.57 & 95.70 & 446.72 \end{bmatrix}.$$

In this case since males are not measured; M_{114} (from (5)) can be written as

$$\sum_{i=1}^{s} (y_i - \bar{y})^2 + \sum_{i=1}^{s} \sum_{j=1}^{d} [(y_{ij} - \bar{y})^2 - (y_{ij} - \bar{y}_1)^2]$$

which is the sum of squares for all females minus the sum of squares for females used as dams. Hence

$$\mathbf{M}_4^{11} = (847.57) \qquad \mathbf{M}_4^{12} = (430.69)$$

$$\mathbf{M}_4^{21} = (430.69) \qquad \mathbf{M}_4^{22} = (551.04)$$

so

$$\mathbf{M}_4 = \begin{bmatrix} \mathbf{M}_4^{11} & \mathbf{M}_4^{12} \\ \mathbf{M}_4^{21} & \mathbf{M}_4^{22} \end{bmatrix} = \begin{bmatrix} 847.57 & 430.69 \\ 430.65 & 551.04 \end{bmatrix},$$

$\mathbf{V}_3$ and $\mathbf{V}_4$ become

$$\mathbf{V}_3 = \mathfrak{G}_P * \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \mathfrak{G}_A * \begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix} + \mathfrak{G}_K * \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

(leaving out the rows and columns associated with sires in $f_3(\boldsymbol{\theta}_i)$) and $\mathbf{V}_4 = \mathfrak{G}_P * (1) + \mathfrak{G}_A * (0) + \mathfrak{G}_K * (0)$.

If we let

$$\mathfrak{G}_P = \begin{bmatrix} \sigma_P^{11} & \sigma_P^{21} \\ \sigma_P^{21} & \sigma_P^{22} \end{bmatrix} \quad \text{and} \quad \mathfrak{G}_A = \begin{bmatrix} \sigma_A^{11} & \sigma_A^{21} \\ \sigma_A^{21} & \sigma_A^{22} \end{bmatrix}$$

we find

$$\mathbf{V}_2 = \begin{bmatrix} \sigma_P^{11} & (1/2)\sigma_A^{11} & \sigma_P^{21} & (1/2)\sigma_A^{21} \\ (1/2)\sigma_A^{11} & \sigma_P^{11} & (1/2)\sigma_A^{21} & \sigma_P^{21} \\ \sigma_P^{21} & (1/2)\sigma_A^{21} & \sigma_P^{22} & (1/2)\sigma_A^{22} \\ (1/2)\sigma_A^{21} & \sigma_P^{21} & (1/2)\sigma_A^{22} & \sigma_P^{22} \end{bmatrix}$$

and

$$\mathbf{V}_4 = \begin{bmatrix} \sigma_P^{11} & \sigma_P^{21} \\ \sigma_P^{21} & \sigma_P^{22} \end{bmatrix}.$$

We note that in this example $\mathfrak{G}_K$ cannot be estimated.

It is convenient to find the information matrix in two stages. We first form $p^2 \times p^2$ matrices $\mathbf{G}_{ij}$ relevant for information on $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$, which are similar to (7). G_{ij} is given by

$$\mathbf{G}_{ij} = (1/2) \sum_{k=1}^{4} \nu_k \operatorname{tr} [(\mathbf{V}_k^{-1} *_{ri} f_k(\boldsymbol{\theta}_i))^{Tr_k} *_{rk} (\mathbf{V}_k^{-1} *_{sk} f_k(\boldsymbol{\theta}_j))]_{p^2} \tag{13}$$

where $*_i$ and $\operatorname{tr}(\ )_i$ represent the direct product of order i and trace of order i used by Thompson [1973]. These generalize the ordinary direct product and trace and essentially work with submatrices of matrices. We need to define another operation of the same type, that is transpose of order i denoted by $(\ )^{T_i}$. If $\mathbf{A}$ is a $mi \times ni$ matrix, then $(\mathbf{A})^{T_i}$ is a $ni \times mi$ matrix with the (j, k)th submatrix of $(\mathbf{A})^{T_i}$ of size $i \times i$ equal to the (k, j)th submatrix of A. If $i = 1$ we have the usual definition of transpose. For instance,

$$(\mathbf{M}_3)^{T_2} = \begin{bmatrix} \mathbf{M}_3^{11} & \mathbf{M}_3^{12} \\ \mathbf{M}_3^{21} & \mathbf{M}_3^{22} \end{bmatrix}$$

i.e., we have interchanged $\mathbf{M}_3^{12}$ and $\mathbf{M}_3^{21}$.

We now derive some of the terms in (13) to show the operations involved. We use as estimates of $\mathfrak{G}_P$ and $\mathfrak{G}_A$

$$\mathfrak{G}_P = \begin{bmatrix} 1.024 & 0.0 \\ 0.0 & 0.999 \end{bmatrix} \quad \text{and} \quad \mathfrak{G}_A = \begin{bmatrix} 0.430 & 0.0 \\ 0.0 & 0.428 \end{bmatrix}$$

this corresponds to the initial estimates used by Thompson [1973] in our parameterization. His $\varrho + \Sigma$ and 2ϱ correspond to our $\mathfrak{d}_P$ and $\mathfrak{d}_A$ so

$$\mathbf{V}_3 = \begin{bmatrix} 1.024 & 0.215 & 0.0 & 0.0 \\ 0.215 & 1.024 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.999 & 0.214 \\ 0.0 & 0.0 & 0.214 & 0.999 \end{bmatrix}$$

and

$$\mathbf{V}_3^{-1} = \begin{bmatrix} 1.0216 & -0.2145 & 0.0 & 0.0 \\ -0.2145 & 1.0216 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0491 & -0.2247 \\ 0.0 & 0.0 & -0.2247 & 1.0491 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

where the $\mathbf{A}$'s are 2×2 matrices defined to show the structure of $\mathbf{V}_3^{-1}{}_2{}^*f_3(\boldsymbol{\theta}_i)$.

$$\mathbf{V}_3^{-1}{}_2^*f_3(\boldsymbol{\theta}_1) = \begin{bmatrix} 1.0216 & -0.2145 & 0.0 & 0.0 \\ -0.2145 & 1.0216 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0491 & -0.2247 \\ 0.0 & 0.0 & -0.2247 & 1.0491 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{A}_{11}f_3(\boldsymbol{\theta}_1) & \mathbf{A}_{12}f_3(\boldsymbol{\theta}_1) \\ \mathbf{A}_{21}f_3(\boldsymbol{\theta}_1) & \mathbf{A}_{22}f_3(\boldsymbol{\theta}_1) \end{bmatrix},$$

and

$$\mathbf{V}_3^{-1}{}_2^*f_3(\boldsymbol{\theta}_2) = \begin{bmatrix} -0.1072 & 0.5108 & 0.0 & 0.0 \\ 0.5108 & -0.1072 & 0.0 & 0.0 \\ 0.0 & 0.0 & -0.1124 & 0.5246 \\ 0.0 & 0.0 & 0.5246 & -0.1124 \end{bmatrix}.$$

In this case, $(\mathbf{V}_3{}_2^*f_3(\boldsymbol{\theta}_i))^{T2} = (\mathbf{V}_3{}_2^*f_3(\boldsymbol{\theta}_i))$ because the submatrices interchanged are the same, but in general this is not so.

$((\mathbf{V}_3^{-1}{}_2^*f_3(\boldsymbol{\theta}_1))^{T2}{}_2^*(\mathbf{V}_3^{-1}{}_2^*f(\boldsymbol{\theta}_2)))$

$$= \begin{bmatrix} -0.2191 & 0.5448 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.5448 & -0.2191 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & -0.2274 & 0.5600 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5600 & -0.2274 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & -0.2272 & 0.5600 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.5600 & -0.2272 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & -0.2358 & 0.5756 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5756 & -0.2358 \end{bmatrix}$$

where

$$\begin{pmatrix} 1.0216 & -0.2145 \\ -0.2145 & 1.0216 \end{pmatrix}\begin{pmatrix} -0.1072 & 0.5108 \\ 0.5108 & -0.1072 \end{pmatrix} = \begin{pmatrix} -0.2191 & 0.5448 \\ 0.5448 & -0.2191 \end{pmatrix}$$

so

$$\text{tr}\,[(\mathbf{V}_3^{-1} *_2 f_3(\boldsymbol{\theta}_1))^{T2} *_2 (\mathbf{V}_3^{-1} *_2 f_3(\boldsymbol{\theta}_2)]_4 = \begin{pmatrix} -0.4382 & 0.0 & 0.0 & 0.0 \\ 0.0 & -0.4548 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.4544 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.4716 \end{pmatrix},$$

since

$$\text{tr}\begin{pmatrix} -0.2191 & 0.5448 \\ 0.5448 & -0.2191 \end{pmatrix} = -0.4382, \quad \text{etc.}$$

In a similar way the other terms of (13) can be calculated.

Formulae (8) and (9) are easily modified in a similar manner. We simply replace the scalar product (.) by the direct product (*), and think of the M^* terms as $p \times p$ matrices rather than scalars. The results are $p^2 \times p^2$ matrices $\mathbf{G}_{ij1}$ and $\mathbf{G}_{ij2}$.

For example, in the numerical example we find

$$\mathbf{M}_{223} = \begin{pmatrix} 175.12 & 92.57 \\ 92.57 & 446.72 \end{pmatrix}$$

and

$$\begin{aligned} \mathbf{M}_2^* &= (\boldsymbol{\delta}_P)^{-1}(\mathbf{M}_{223} - 500\boldsymbol{\delta}_P)(\boldsymbol{\delta}_P)^{-1} \\ &= \begin{pmatrix} 1.024 & 0.0 \\ 0.0 & 0.999 \end{pmatrix}^{-1}\left[\begin{pmatrix} 175.12 & 92.57 \\ 92.57 & 446.72 \end{pmatrix} - \begin{pmatrix} 512.0 & 0.0 \\ 0.0 & 499.5 \end{pmatrix}\right]\begin{pmatrix} 1.024 & 0.0 \\ 0.0 & 0.999 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} -321.2738 & 90.4909 \\ 90.4909 & -52.8857 \end{pmatrix}. \end{aligned}$$

(In the 1973 paper there is a miscalculation, the -52.6252 term in $\mathbf{M}_{11}{}^*$ should be -52.8857).

The other $\mathbf{M}^*$ terms are null or do not exist because only females are measured; $\mathbf{V}_3{}^*$ (from (9)) is

$$\begin{aligned} \mathbf{V}_3{}^* &= \boldsymbol{\delta}_P - (1/4)\boldsymbol{\delta}_A(\boldsymbol{\delta}_P)^{-1}\boldsymbol{\delta}_A \\ &= \begin{pmatrix} 0.9789 & 0.0 \\ 0.0 & 0.9532 \end{pmatrix}. \end{aligned}$$

$\mathbf{G}_{122}$ is (from (8))

$$\begin{aligned} \mathbf{G}_{122} &= (-1/4)\mathbf{M}_2{}^* * (\boldsymbol{\delta}_P)^{-1}\boldsymbol{\delta}_A\mathbf{V}_3{}^* \\ &= \begin{pmatrix} 34.4545 & -9.7045 & 0.0 & 0.0 \\ -9.7045 & 5.6716 & 0.0 & 0.0 \\ 0.0 & 0.0 & 36.1002 & -10.1681 \\ 0.0 & 0.0 & -10.1681 & 5.9625 \end{pmatrix}. \end{aligned}$$

The second stage is to form the information matrix (**H**) from the **G**'s. This follows the scheme discussed in Thompson [1973].

The matrices $\boldsymbol{\theta}_i$ are symmetric, so there are only $(p(p+1)/2)$ distinct elements in $\boldsymbol{\theta}_i$. We represent these by $(\boldsymbol{\theta}_i)_T$, a column vector formed by taking the elements of the lower triangle of $\boldsymbol{\theta}_i$; that is, the $(j(j-1)\ 2+k)$th element of $(\boldsymbol{\theta}_i)$ contains the (j, k)th element of $\boldsymbol{\theta}_i$.

Associated with $(\boldsymbol{\theta}_i)_T$ and $(\boldsymbol{\theta}_j)_T$ is a submatrix $\mathbf{H}_{ij}$ of $\mathbf{H}$ given by

$$\mathbf{J}(\mathbf{G}_{ij} + \mathbf{G}_{ji1} + \mathbf{G}_{ji2})\mathbf{J}', \tag{14}$$

where $\mathbf{J}$ is the Jacobian of the transformation from $(\boldsymbol{\theta}_i)_R$ to $(\boldsymbol{\theta}_i)_T$ and $(\boldsymbol{\theta}_i)_R$ represents a column vector with its $((k-1)p+j)$th element containing the (j, k)th element of $(\boldsymbol{\theta}_i)$.

When sires are not measured, $d = 1$, $n = 1$ and $K = 0$ (14) are consistent with formulae given by Thompson [1973]. In that paper, $\boldsymbol{\varrho} + \boldsymbol{\Sigma}$ and $2\boldsymbol{\varrho}$ are equivalent to $\boldsymbol{\sigma}_P$ and $\boldsymbol{\sigma}_A$ in this paper.

In the univariate case, it seems reasonable to use the variance of the estimate of heritability as a criteria to judge designs. In the multivariate case, it does not seem so simple. For two variates one could, for example, choose designs on the basis of the variance of the estimate of genetic correlation. Tallis [1959], for instance, used this as a criterion in considering designs on half-sibs. But it seems a moot point to me that this is necessarily the most important parameter. The uses that are made of the estimates should influence the choice of criteria that are used to classify the designs. For instance, with two variates one might be interested in asking if using the additional information from a second variate would improve a selection program based on one variate, or alternatively, in the case of indirect selection, if selecting on a second variate would be preferable to selecting on the trait of interest.

Given a legitimate criterion to judge a design, various possibilities are open. Reeve [1955] suggested that one might run two experiments in parallel to estimate a genetic correlation, one half selecting and mating assortatively parents on one trait and the other half selecting and mating assortatively on the other variate. Another alternative is to carry out elliptical truncation (Tallis [1963]). In this case, animals with the largest values of $(\mathbf{x}-\boldsymbol{\mu})'\mathbf{A}(\mathbf{x}-\boldsymbol{\mu})$ are used as parents, where $\mathbf{x}$ is the $p \times 1$ vector of records, $\boldsymbol{\mu}$ is the $p \times 1$ vector of means, and $\mathbf{A}$ is a $p \times p$ symmetric positive-definite matrix. When $p = 1$ this corresponds to picking the animals with the largest and smallest records. This strategy, with $\mathbf{A} = \boldsymbol{\sigma}_P^{-1}$, was evaluated over a limited number of cases. It gave reductions of the order of 10 to 20 percent for the variance of genetic correlation when compared with Reeve's procedure.

Acknowledgments

I am grateful for the comments of Professor Oscar Kempthorne.

Most of this work was done in connection with research under the NIH Grant 13827. Part was done under Research Grant NSF GP-2614.

Plan d'expériences Pour une Estimation d'héritabilité quand des Observations Sont Disponibles sur Parents et Descendants

Résumé

On considère le plan d'expériences pour estimer l'héritabilité lorsque les données proviennent de parents et de descendants, les données sur descendance ayant une structure hiérarchique

L'estimation du maximum de vraisemblance univariate est discutée et des extensions au cas multivariate sont esquissées. L'efficacité de l'estimateur du maximum de vraisemblance est évaluée dans des cas où des estimateurs par régression simple sont disponibles. Des plans optimaux pour l'estimation du maximum de vraisemblance sont donnés quand différentes stratégies de sélection et de croisement sont suivies. La variance de l'estimateur d'héritabilité peut être approximativement diminuée de moitié relativement aux plans dans lesquels aucune sélection des parents n'est faite.

References

Curnow, R. N. [1961]. The estimation of repeatability and heritability from records subject to culling. *Biometrics 17*, 553–66.

Henderson, C. R., Kempthorne, O., Searle, S. R., and von Krosigk, C. M. [1959]. The estimation of environmental and genetic trends from records subject to culling. *Biometrics 15*, 192–218.

Hill, W. G. [1970]. Design of experiments to estimate heritability by regression of offspring on selected parents. *Biometrics 26*, 566–71.

Hill, W. G. and Nicholas, F. W. [1974]. Estimation of heritability by both regression of offspring on parent and intra-class correlation of sibs in one experiment. *Biometrics 30*, 447–68.

Latter, B. D. H. and Robertson, A. [1960]. Experimental design in the estimation of heritability by regression methods. *Biometrics 16*, 348–53.

Patterson, H. D. and Thompson, R. [1971]. Recovery of inter-block information when block sizes are unequal. *Biometrika 58*, 545–54.

Reeve, E. C. R. [1955]. The variance of the genetic correlation coefficient. *Biometrics 11*, 357–74.

Robertson, A. [1959]. Experimental design in the evaluation of genetic parameters. *Biometrics 15*, 219–26.

Searle, S. R. [1966]. *Matrix Algebra for the Biological Sciences*. Wiley, New York.

Tallis, G. M. [1959]. Sampling errors of genetic correlation coefficients calculated from analyses of variance and covariance. *Aust. J. of Statist. 1*, 35–43.

Tallis, G. M. [1963]. Elliptical and radial truncation in normal populations. *Ann. Math. Statist. 34*, 940–4.

Thompson, R. [1973]. The estimation of variance and covariance components with an application when records are subject to culling. *Biometrics 29*, 527–50.

Received October 1974, *Revised July* 1975

Appendix

In this appendix we derive the information matrix G^* (given by (4), (5) and (6) in Section 2) when there is selection and assortative mating of the parents.

The log likelihood (L) maximized is (from (6))

$$\text{const} - (1/2) \sum_{k=1}^{4} [\nu_k \ln |\mathbf{V}_k| + \text{tr}\,(\mathbf{M}_k \mathbf{V}_k^{-1})]$$

so

$$\frac{\partial L}{\partial \theta_i} = -(1/2) \sum_{k=1}^{4} \left[\nu_k \,\text{tr} \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_i} \right] - \text{tr} \left[M_k V_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_i} \mathbf{V}_k^{-1} \right] \right]$$

and

$$\frac{\partial^2 L}{\partial \theta_i \, \partial \theta_j} = +(1/2) \sum_{k=1}^{4} \left[\nu_k \,\text{tr} \left[\mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_i} \mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_j} \right] - 2 \,\text{tr} \left[\mathbf{M}_k \mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_i} \mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_j} \mathbf{V}_k^{-1} \right] \right] \qquad \text{(A1)}$$

(using the fact that

$$\text{tr}\left[\mathbf{M}_k\mathbf{V}_k^{-1}\frac{\partial\mathbf{V}_k}{\partial\theta_i}\mathbf{V}_k^{-1}\frac{\partial\mathbf{V}_k}{\partial\theta_j}\mathbf{V}_k^{-1}\right] = \text{tr}\left[\mathbf{M}_k\mathbf{V}_k^{-1}\frac{\partial\mathbf{V}_k}{\partial\theta_j}\mathbf{V}_k^{-1}\frac{\partial\mathbf{V}_k}{\partial\theta_i}\mathbf{V}_k^{-1}\right]$$

since $\mathbf{M}_k$ and $\mathbf{V}_k$ are symmetric).

We require

$$E\left[\frac{\partial^2 L}{\partial\theta_i\,\partial\theta_j}\right]$$

and evaluate it in three stages.

a) $k = 1$

The expected value of $\mathbf{M}_1$ is $\nu_1\mathbf{V}_1$ and independent of the selection scheme used. The contribution to

$$E\left[\frac{\partial^2 L}{\partial\theta_i\,\partial\theta_j}\right]$$

is

$$-(\nu_1/2)\,\text{tr}\,(\mathbf{V}_1^{-1}f_1(\theta_i)\cdot\mathbf{V}_1^{-1}f_1(\theta_j)). \tag{A2}$$

b) $k = 2$ *and* 3

For these it is convenient to partition $\mathbf{M}_2$ and $\mathbf{M}_3$ into four parts relating to measurements on parents and offspring. We express the expected values of $\mathbf{M}_2$ and $\mathbf{M}_3$ conditional on the parental values. The expected values are of the form

$$\begin{pmatrix}\nu(\mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}) + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{M}_{22}\mathbf{V}_{22}^{-1}\mathbf{V}_{21} & \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{M}_{22}\\ \mathbf{M}_{22}\mathbf{V}_{22}^{-1}\mathbf{V}_{21} & \mathbf{M}_{22}\end{pmatrix} \tag{A3}$$

where

$$\mathbf{M}_{11} = M_{112}\,, \qquad \mathbf{M}_{12} = M_{122}\,,$$
$$\mathbf{M}_{21} = M_{212}\,, \qquad \mathbf{M}_{22} = M_{222}\,,$$

and

$$\nu = \nu_2 \quad \text{for} \quad E(\mathbf{M}_2)$$

and

$$\mathbf{M}_{11} = M_{113}\,, \qquad \mathbf{M}_{12} = (M_{123} \quad M_{123}),$$
$$\mathbf{M}_{21} = \begin{pmatrix}M_{213}\\ M_{313}\end{pmatrix} \qquad \mathbf{M}_{22} = \begin{pmatrix}M_{223} & M_{233}\\ M_{323} & M_{333}\end{pmatrix}$$
$$\nu = \nu_3 \quad \text{for} \quad E(\mathbf{M}_3),$$

with similar partitions for $\mathbf{V}_2$ and $\mathbf{V}_3$.

If $E(\mathbf{M})$ is of the form (A3) we find, using the identity

$$\mathbf{V}^{-1} = \begin{pmatrix}(\mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21})^{-1} & -(\mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21})^{-1}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\\ -\mathbf{V}_{22}^{-1}\mathbf{V}_{21}(\mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{21}^{-1}\mathbf{V}_{21})^{-1} & \mathbf{V}_{22}^{-1} + \mathbf{V}_{22}^{-1}\mathbf{V}_{21}(\mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21})^{-1}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\end{pmatrix}$$

that

$$E(\mathbf{V}^{-1}\mathbf{M}\mathbf{V}^{-1}) = \nu\mathbf{V}^{-1} + \begin{pmatrix}0 & 0\\ 0 & \mathbf{V}_{22}^{-1}(\mathbf{M}_{22} - \nu\mathbf{V}_{22})\mathbf{V}_{22}^{-1}\end{pmatrix}$$

so that

$$(1/2)E\left(\nu \operatorname{tr}\left(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_i}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\right) - 2\operatorname{tr}\left(\mathbf{M}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_i}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{V}^{-1}\right)\right.$$

$$= -(\nu/2)\operatorname{tr}\left[\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_i}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\right] - \operatorname{tr}\left[\begin{pmatrix}0 & 0\\ 0 & \mathbf{V}_{22}^{-1}(\mathbf{M}_{22} - \nu\mathbf{V}_{22})\mathbf{V}_{22}^{-1}\end{pmatrix}\frac{\partial \mathbf{V}}{\partial \theta_i}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\right].$$

Again using the identity for $\mathbf{V}^{-1}$ we can simplify this to

$$-(\nu/2)\operatorname{tr}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_i}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j} - \operatorname{tr}\left[(\mathbf{M}_{22} - \nu\mathbf{V}_{22})\mathbf{V}_{22}^{-1}\frac{\partial \mathbf{V}_{22}}{\partial \theta_i}\mathbf{V}_{22}^{-1}\frac{\partial \mathbf{V}_{22}}{\partial \theta_j}\mathbf{V}_{22}^{-1}\right]$$

$$- \operatorname{tr}\left[(\mathbf{M}_{22} - \nu\mathbf{V}_{22})\mathbf{V}_{22}^{-1}\left[\frac{\partial \mathbf{V}_{21}}{\partial \theta_i} - \frac{\partial \mathbf{V}_{22}}{\partial \theta_i}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\right](\mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21})^{-1}\right.$$

$$\left.\cdot\left[\frac{\partial \mathbf{V}_{12}}{\partial \theta_j} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\frac{\partial \mathbf{V}_{22}}{\partial \theta_j}\right]\mathbf{V}_{22}^{-1}\right]$$

$$= (-\nu/2)\operatorname{tr}\left[\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_i}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\right] - \operatorname{tr}\left[(\mathbf{M}_{22} - \nu\mathbf{V}_{22})\mathbf{V}_{22}^{-1}\frac{\partial \mathbf{V}_{22}}{\partial \theta_i}\mathbf{V}_{22}^{-1}\frac{\partial \mathbf{V}_{22}}{\partial \theta_j}\mathbf{V}_{22}^{-1}\right]$$

$$- \operatorname{tr}\left[(\mathbf{M}_{22} - \nu\mathbf{V}_{22})\frac{\partial(\mathbf{V}_{22}^{-1}\mathbf{V}_{21})}{\partial \theta_i}(\mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21})^{-1}\frac{\partial(\mathbf{V}_{12}\mathbf{V}_{22}^{-1})}{\partial \theta_j}\right].$$

When

$$k = 2; \qquad \mathbf{V}_{22}^{-1}\mathbf{V}_{21} = \frac{n}{2}(\sigma_P^2)^{-1}\sigma_A^2$$

so we can write

$$\frac{\partial(\mathbf{V}_{22}^{-1}\mathbf{V}_{21})}{\partial \theta_i} = (-1)^i\frac{n}{2}(\sigma_P^2)^{-1}\theta_i\sigma_A^2 \qquad (i = 1, 2)$$

and $\mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21} = V_2^*$ (from (9)).

Hence the contribution from $k = 2$ to

$$E\left[\frac{\partial^2 L}{\partial \theta_i\,\partial \theta_i}\right]$$

is

$$-(\nu_2/2)\operatorname{tr}\left[\mathbf{V}_2^{-1}f_2(\theta_i)\cdot\mathbf{V}_2^{-1}f_2(\theta_j)\right] - M_1^*\frac{\partial \sigma_P^2}{\partial \theta_i}(\sigma_P^2)^{-1}\frac{\partial \sigma_P^2}{\partial \theta_j} - G_{ij1}. \tag{A4}$$

Similarly for $k = 3$ the contribution is

$$-(\nu_3/2)\operatorname{tr}\left[\mathbf{V}_3^{-1}f_3(\theta_i)\cdot\mathbf{V}_3^{-1}f_3(\theta_j)\right] - ((M_2^*/d) + M_3^*)\frac{\partial \sigma_P^2}{\partial \theta_i}(\sigma_P^2)^{-1}\frac{\partial \sigma_P^2}{\partial \theta_j} - G_{ij2}. \tag{A5}$$

c) $k = 4$

$\mathbf{M}_4$ is the sum of the sums of squares between females and between males in the parent generation minus the sum of squares between females with offspring and minus the sum of squares between males with offspring.

$$E(M_4) = ((D' - 1) + (s' - 1))\mathbf{V}_4 - (M_{222} + (M_{223}/d)) - M_{333}$$

$$= \nu_4\mathbf{V}_4 - (M_{222} - d_2\sigma_P^2) - (M_{223} - d_3 d\sigma_P^2)/d - (M_{333} - d_3).$$

The contribution to

$$E\left[\frac{\partial^2 L}{\partial\theta_i\,\partial\theta_j}\right]$$

from $k = 4$ is

$$-(\nu_4/2)\,\mathrm{tr}\,[V_4^{-1}f_4(\theta_i)V_4^{-1}f_4(\theta_j)] + (M_1^* + M_2^*/d + M_3^*)\frac{\partial \mathbf{V}_4}{\partial\theta_i}\,\mathbf{V}_4^{-1}\frac{\partial \mathbf{V}_4}{\partial\theta_j}. \qquad \text{(A6)}$$

Combining (*A*2), (*A*4), (*A*5), and (*A*6) we find that

$$-E\left[\frac{\partial^2 L}{\partial\theta_i\,\partial\theta_j}\right] \quad \text{is} \quad G_{ij}^*.$$

16

Design of experiments to estimate parent-offspring regression using selected parents

by

W.G. Hill and R. Thompson

Anim. Prod. 1977, 24: 163–168

DESIGN OF EXPERIMENTS TO ESTIMATE OFFSPRING-PARENT REGRESSION USING SELECTED PARENTS

W. G. HILL
Institute of Animal Genetics, Edinburgh EH9 3JN

AND

R. THOMPSON
ARC Unit of Statistics, Edinburgh EH9 3JZ

SUMMARY

Heritability or the genetic regression of one trait on another, for example live-weight gain in field *v.* performance test station conditions, can be estimated efficiently by rearing offspring only from high- and low-ranking individuals, and estimating the regression of offspring on parent performance. Formulae are given for the optimum proportion of available animals to select as parents when the relative numbers of individuals in the parent and progeny generation are fixed. Although the optimum depends on the value of the parameter to be estimated, the designs are very robust to poor estimates.

INTRODUCTION

IF heritability of a trait is estimated by the regression of progeny on parent performance, the parents used can be selected for this trait. The estimates should not then be appreciably biased unless there are loci, particularly with non-additive gene action, which have a large effect on the trait (Reeve, 1961), or the genetic and environmental distributions are skewed to different extents (Nishida and Abe, 1974). If only the highest and lowest ranking parents are selected, the variance of the independent variate is increased and the sampling variance of the regression estimate of heritability can be reduced. For the same total number of individuals measured, this high and low selection of parents gives heritability estimates with sampling variance about one-half of estimates obtained using all available as parents (Hill, 1970) and the method is much more efficient than using regression solely on high-selected parents since the variance within the high group alone is small. With high-low selection, regression estimates of heritability are more efficient than half-sib correlation estimates, which also may be biased if any selection among parents is practised.

Use could be made of regression estimation on high- and low-selected parents in cattle, for example to estimate the heritability of growth rate in performance testing stations from fathers and sons. Alternatively, the parental generation might be reared in test facilities and the progeny of selected individuals reared under field conditions to evaluate a central performance testing station. Considering growth rate in the two environments as different traits, the design allows for estimation of the genetic regression of one trait on another. In this way, J. B. Owen (personal com-

munication) is evaluating a performance test using early weaning and artificial rearing in sheep.

The proportion of the parental generation which is selected as breeders and the number of progeny from each selected individual can be varied. Hill (1970) gave optimum designs for estimates of heritability obtained by regression and Thompson (1976) for estimates of heritability obtained by combining regression and sib covariance by maximum likelihood. In these studies the variance of the heritability estimate was minimized for a given total number of individuals over the two generations. In practice, however, the total number of individuals which can be recorded in each of the two generations may be fixed, so an increase in the number of individuals chosen as parents leads to a corresponding proportionate decrease in the number of progeny in each family. For example, to make full use of expensive test facilities, the same number of individuals would be recorded in each generation, whereas in a field test of the progeny, many more progeny might be reared. In this paper the optimum designs are derived for such experiments in which the relative numbers (R) in the parental and progeny generations are fixed; the absolute numbers of each affect the variance of the estimate, but not the optimum proportion of potential parents selected.

ANALYSIS

In the parental generation, assume a total of M unrelated individuals of one sex are recorded. Of these a fraction p with the highest score and p with the lowest score are chosen and a total of RM progeny reared and recorded. Thus there are $n = R/2p$ progeny per family, assumed to be the same for each family, and a total of $M(1+R)$ individuals are recorded over the two generations. Let:

σ_Y^2, σ_Z^2 be the variances of measurements, assumed normally distributed, on parents and progeny, respectively,

b be the regression of progeny on parent performance and $\hat{b}$ its estimate,

r be the correlation of progeny and parent performance, and

t the intra-class correlation of family members,

where σ_Y^2, σ_Z^2, r and t refer to unselected random mating populations and the regression, b, is assumed to be unaffected by selection.

The sum of squares among selected parents, assumed of sufficient numbers that infinite rather than order statistic theory can be used, equals $2Mp(1+ix)\sigma_Y^2$ where x and i are the abscissa and selection differential on the standardized normal curve corresponding to p (Hill, 1970). The variance about regression of the mean of a family of size n equals $[t+(1-t)/n-r^2]\sigma_Z^2$ and thus

$$V(\hat{b}) = \frac{[1-t+n(t-r^2)]\sigma_Z^2}{2Mnp(1+ix)\sigma_Y^2} = \frac{[2p(1-t)+R(t-r^2)]\sigma_Z^2}{2pMR(1+ix)\sigma_Y^2} \qquad (1)$$

When very few animals are selected, the variance is likely to be slightly greater than in (1).

The optimum proportion to select for minimizing $V(\hat{b})$ is found by differentiation of (1) with respect to p. The solution is given by

$$\frac{2(1-t)}{R(t-r^2)} = \frac{x^2}{p(1+ix-x^2)} = W(p) \qquad (2)$$

say, which can also be obtained by appropriate combination of the partial derivatives with respect to p and n given by Hill (1970).

Figure 1 shows a plot of W(p) against p. The corresponding results when n and p are both optimized are tabulated by Hill (1970) and Thompson (1976). A plot of $1+ix$ against p is also given in Figure 1 for substitution into (1) to compute variances.

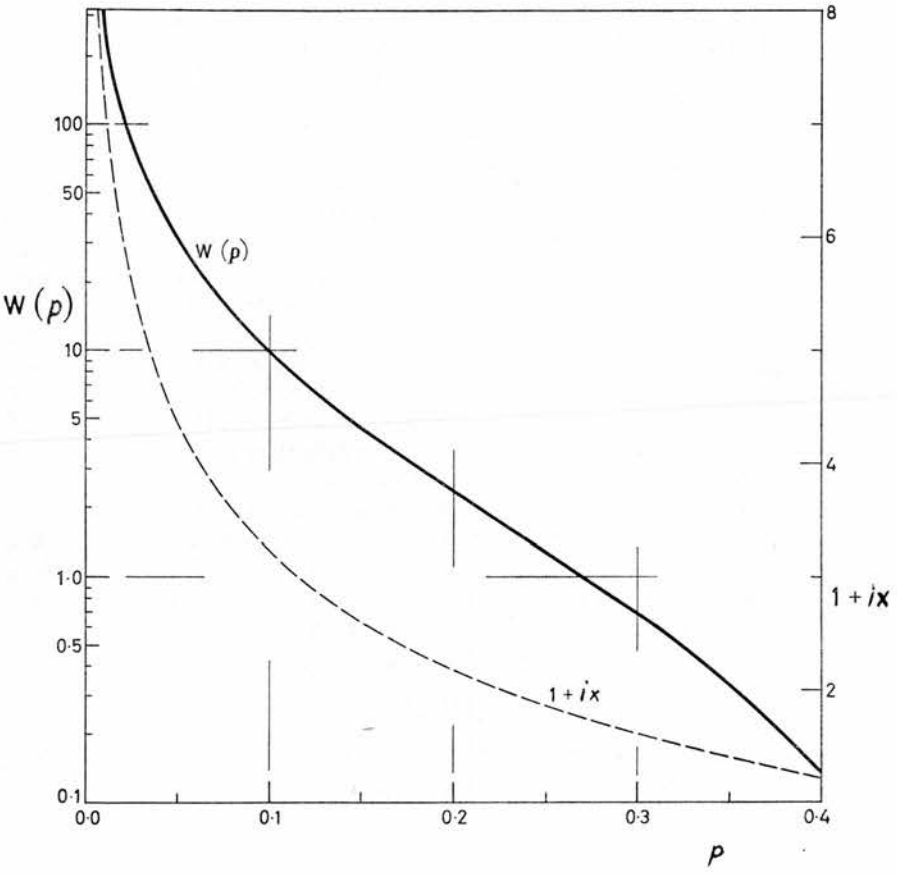


FIG. 1. Values of W(p) and $1+ix$ plotted against p.

EXAMPLES

In cattle, assume that potential sires are recorded for growth rate in a testing station and the best and worst are selected and mated to a random group of dams with progeny in half-sib families. If the same trait is measured on progeny in the same facilities to obtain an estimate of heritability (h^2) and there is no environmental covariance of sibs, $\sigma^2_Y = \sigma^2_Z$, $b = r = h^2/2$ and $t = h^2/4$. Substituting in (1)

$$V(\hat{h}^2) = 4V(\hat{b}) = \frac{2p(4-h^2)+Rh^2(1-h^2)}{2pMR(1+ix)}. \qquad (3)$$

and substituting in (2) to find the optimum,

$$W(p) = \frac{2(4-h^2)}{Rh^2(1-h^2)}. \quad (4)$$

Assuming the same numbers are recorded each generation, $R = 1$, and taking a typical value of $h^2 = 0{\cdot}4$ for growth rate, from (4), $W(p) = 30$ and from Figure 1, $p = 0{\cdot}05$ approximately, and $1+ix = 4{\cdot}4$. Hence, from (3) $V(\hat{h}^2) = 1{\cdot}36/M$, so with facilities for $M = 100$ animals, $SE(\hat{h}^2) = 0{\cdot}12$ approximately. Alternatively, if field performance of progeny were to be evaluated on 16 times as many animals as in the station, $R = 16$ and, assuming $h^2 = 0{\cdot}4$ but the genetic correlation is 0·75 between test station and field performance, $t = h^2/4 = 0{\cdot}1$ and $r = 0{\cdot}75h^2/2 = 0{\cdot}15$, giving $W(p) = 1{\cdot}45$ from (2) and $p = 0{\cdot}24$ from Figure 1.

There are other designs in which regression on a single parent is computed and for simplicity these are illustrated for heritability estimation. If progeny are in full sib families, and heritability of, for example, litter size in mice is estimated by daughter on dam regression, then $b = r = t = h^2/2$. For estimation of the heritability of, say, milk yield in cattle by daughter on dam regression in half-sib families, $b = r = h^2/2$ and $t = h^2/4$, but the variance within families is also reduced by $h^2/4$ and the numerator of (1) becomes $2p(1-h^2/2)+Rh^2(1-h^2)/4$. When parents and progeny are in different environments, terms such as $h^2/2$ and r have to be replaced by one-half of the square root of the product of the heritabilities in the two environments and the genetic correlation between them.

The formulae can be adapted for regression on mid-parent value in full sib families, for example with growth rate in pigs or mice. Assume there are records on $M/2$ males and $M/2$ females, so with a total of RM progeny, the family size is R/p. If the selected parents are mated with perfect positive assortment to maximize the sum of squares, from (1)

$$V(\hat{b}) = \frac{[p(1-t)+R(t-r^2)]\sigma_Z^2}{pMR(1+ix)\sigma_Y^2}.$$

At the optimum $W(p) = (1-t)/[R(t-r^2)]$ and Figure 1 can be used. For heritability estimation, $b = h^2$, $t = h^2/2$, $r = h^2/\sqrt{2}$ and $W(p)$ reduces to $(2-h^2)/[Rh^2(1-h^2)]$.

DISCUSSION

Efficiency. Let us compare the efficiency of these designs using high-low selection with alternative regression estimators. In the example of heritability estimation using son on sire regression, the variance obtained was $V(\hat{h}^2) = 1{\cdot}36/M$ with a true heritability of 0·4 and $R = 1$. By contrast, if no selection were practised and one progeny per sire were taken, then $p = 0{\cdot}5$ and $x = 0$ in (1) or (3) and $V(\hat{h}^2) = 3{\cdot}84/M$, i.e. 2·8 times greater. If fewer parents were randomly sampled to increase the family size to n, the variance would be greater, $(3{\cdot}6+0{\cdot}24n)/M$ or $4{\cdot}8/M$ with $n = 5$. A disadvantage of selecting high and low, or of random sampling, is that inferior animals have to be used as parents. If only high-scoring parents are chosen, the variance among them is reduced to $[1-i(i-x)]\sigma^2$ and the regression estimator is less efficient. For example, if all individuals scoring above

average were used as parents for the heritability estimation, the predicted variance is $11{\cdot}2/M$.

Robustness. Since h^2, t and r, are not known *a priori*, designs should be robust to poor estimates of these parameters, and it turns out that those involving selection usually are. For illustration, variances of heritability estimates from progeny on sire regression (3) were computed for a range of h^2 values at fixed and at optimum values of p and given R. These are

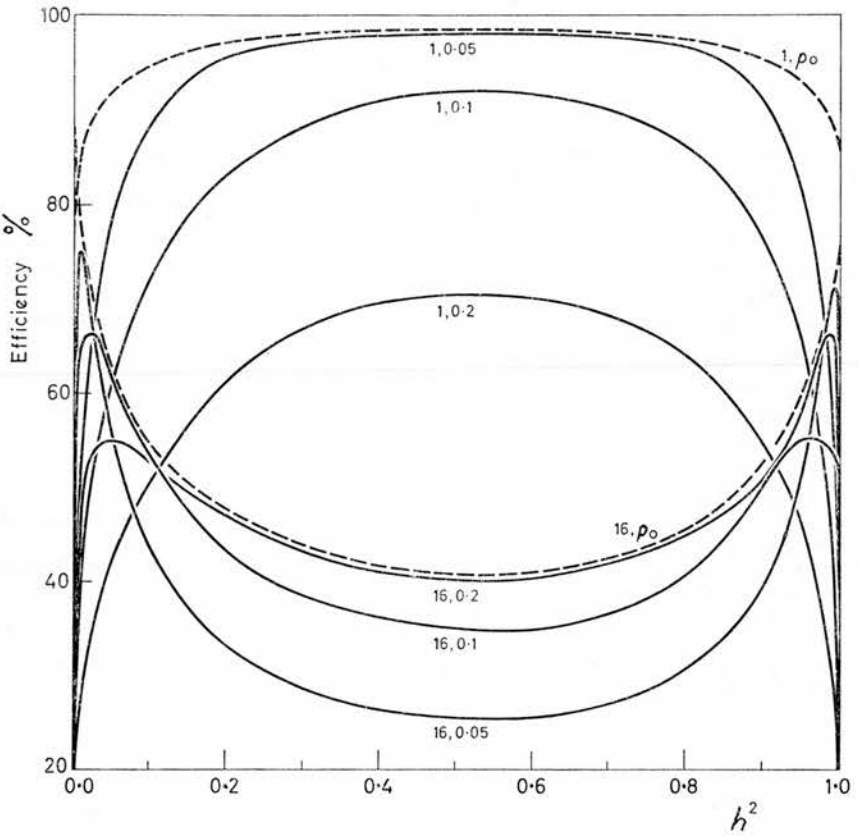


FIG. 2. Efficiency of alternative estimates of heritability obtained by progeny on sire regression, expressed as the variance for a fixed total number recorded relative to that when both p and R are optimized. Results are given for specified p, R with p fixed or optimal (p_0).

compared with variances obtained when both p and R (or equivalently p and family size) are optimized (Hill, 1970) in Figure 2. In each case the variances were expressed in proportion to the total number recorded over the two generations as $M(1+R)\mathrm{V}(\hat{h}^2)$. The efficiency when both p and R are optimized is shown as 100% and the variances of other estimators expressed relative to this base. Figure 2 shows that, for a wide range of heritability values and fixed R, $p = 0{\cdot}05$ is efficient for $R = 1$ and $p = 0{\cdot}2$ for $R = 16$, and that the design with $p = 0{\cdot}05$ and $R = 1$ is not much less efficient than if both p and R are optimized.

Combination of regression and intra-class correlation. When the same trait is measured on parents and progeny and there is no environmental or dominance covariance of sibs, maximum likelihood (ML) can be used to give a best estimate of heritability combining the covariances between progeny and parent and between sibs, even though selection is practised (Thompson, 1976). In Thompson's analysis both R and p were allowed to vary in order to find the optimum design; the optima have now been computed for fixed R. Taking the example of data on sire and progeny in half-sib families, when combining information by ML the optimum value of p is somewhat lower at low values of h^2 and R than when only the regression is used (at $h^2 = 0{\cdot}1$, $R = 1$, $p = 0{\cdot}021$ rather than $0{\cdot}025$). If, however, the value of p is derived using Figure 1 assuming the use of regression alone, the variance of the ML estimator of heritability is very close to that obtained using the optimum value of p. The benefit of analysing the data by ML is greatest when the heritability is low and R is large, the efficiency being more than doubled at $h^2 = 0{\cdot}05$, $R = 16$. For heritabilities estimated using the other designs mentioned above, selection intensities among parents computed assuming only the regression estimate is used are also very efficient if ML is subsequently used in the analysis.

In conclusion, designs based on (2) using quite poor predictions of the parameters are likely to be efficient over a wide range of true parameter values, whether the results are analysed solely by regression or by maximum likelihood.

REFERENCES

HILL, W. G. 1970. Design of experiments to estimate heritability by regression of offspring on selected parents. *Biometrics* **26**: 566–571.

NISHIDA, A. and ABE, T. 1974. The distribution of genetic and environmental effects and the linearity of heritability. *Can. J. Genet. Cytol.* **16**: 3–10.

REEVE, E. C. R. 1961. A note on non-random mating in progeny tests. *Genet. Res.* **2**: 195–203.

THOMPSON, R. 1976. Design of experiments to estimate heritability when observations are available on parents and offspring. *Biometrics* **32**: 283–304.

(*Received* 1 *March* 1976)

17

Design of multivariate selection experiments to estimate genetic parameters

by

N.D. Cameron and R. Thompson

Theor Appl Genet (1986) 72:466–476

THEORETICAL AND APPLIED GENETICS

© Springer-Verlag 1986

Design of multivariate selection experiments to estimate genetic parameters

N. D. Cameron and R. Thompson

AFRC Animal Breeding Research Organisation, West Mains Road, Edinburgh EH9 3JQ, Scotland

Received July 27, 1985; Accepted December 20, 1985
Communicated by A. Robertson

Summary. The precision of estimates of genetic variances and covariances obtained from multivariate selection experiments of various designs are discussed. The efficiencies of experimental designs are compared using criteria based on a confidence region of the estimated genetic parameters, with estimation using both responses and selection differentials and offspring-parent regression. A good selection criterion is shown to be to select individuals as parents using an index of the sums of squares and crossproducts of the phenotypic measurements. Formulae are given for the optimum selection proportion when the relative numbers of individuals in the parent and progeny generations are fixed or variable. Although the optimum depends on "a priori" knowledge of the genetic parameters to be estimated, the designs are very robust to poor estimates. For bivariate uncorrelated data, the variance of the estimated genetic parameters can be reduced by approximately 0.4 relative to designs of a more conventional nature when half of the individuals are selected on one trait and half on the other trait. There are larger reductions in variances if the traits are correlated.

Key words: Experimental design – Genetic parameter estimation – Multiple traits – Selection – Canonical variates

Introduction

Precise, unbiased estimates of genetic parameters, such as heritability and genetic correlations, are necessary to optimise breeding programs and to predict rates of change for various selection schemes. These parameters can be estimated from the covariance among collateral relatives or from the regression of the progeny performance on that of their parents. Appropriate equations for the variances of these estimates obtained by such methods are well documented (e.g. Falconer 1981). Equations for calculating the variance of heritability estimates derived from single-trait selection experiments for various designs have been derived by Hill (1971).

One experimental design objective in single-trait selection experiments is to minimise the variance of the heritability estimate which is influenced by factors such as population size, selection intensity, family size, the genetic and phenotypic parameters and the number of generations of selection. Using prior information about the parameters of interest, efficient selection experiments can be designed to obtain precise, unbiased estimates of the parameters using the equations of Hill (1971).

When dealing with two or more traits, the genetic variances and covariances are parameters of interest and, as Thompson (1976) has noted, it is not obvious what the optimal design objective should be. Robertson (1959) and Tallis (1959) discussed the sampling variance of the genetic correlation coefficient and suggested that designs which are efficient for heritability estimation are also efficient for estimation of genetic correlations. For two traits, individuals in the parental generation could be split into two groups, selecting high and low within one group for trait X_1 and selecting high and low within the other group for trait X_2 (Reeve 1955) and studying either the regression of offspring traits on traits of the selected parents or the direct and correlated responses to selection. However, this may not be the most efficient design in an overall sense. Indices using both traits could be used as the selection criteria, rather than selecting directly on the traits measured. However, Gunsett et al. (1984) suggest a strong dependency of the design efficiency on the index weights used. We discuss these techniques for estimating genetic variances and covariances for two traits and compare the efficiencies of different selection designs.

We consider, in detail, two generation selection experiments when parental observations are only taken on one sex. A different experimental design to the classical high-low individual selection method is examined and it is shown to be more efficient and robust.

Optimality criteria

Given a regression problem, $Y = X\beta + e$, where Y is a vector of the dependent variable, X is the design matrix for the independent variables and e is the vector of residuals with variance-covariance matrix V, then the confidence ellipsoid of the generalised least squares estimate $\hat{\beta}$ of β, $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y$, with variance $(X'V^{-1}X)^{-1}$, has the form

$$[\beta: (\beta-\hat{\beta})'X'V^{-1}X(\beta-\hat{\beta}) < \text{constant}]$$

for any specified confidence coefficient. The content of the ellipsoid (e.g. volume in three dimensions) is proportional to $|X'V^{-1}X|^{-1/2}$. Therefore one design criterion is to minimise the content of the ellipsoid or to maximise $|X'V^{-1}X|$, the D-optimality criterion (Silvey 1980). The determinant of $X'V^{-1}X$ will be denoted by $DET(\beta)$. The D-optimality criterion has the useful invariance property that if a design X maximises $DET(\beta)$, then the same design X also maximises $DET(T^*\beta)$, where T^* is a full rank transformation matrix. Therefore, a design that is optimal for estimation of β is also optimal for a linear transformation, $T^*\beta$, of β. There are other overall criteria; for example, to maximise the trace of $X'V^{-1}X$ (the sum of the diagonal elements of the matrix) or to maximise the minimum eigenvalue of $X'V^{-1}X$, but these do not have this invariance property.

Standardisation of traits

The genetic and phenotypic variance-covariance matrices for the traits will be denoted by G and P, respectively. We consider cases of standardised traits, with mean zero, when the diagonal elements of the P matrix are equal to one and assume that the traits are normally distributed. The methods and designs considered can be applied to multivariate data but are developed using bivariate data. The genetic variances and covariances of the standardised traits are then heritabilities (h_1^2 and h_2^2) and co-heritabilities ($r_A h_1 h_2$ where r_A is the genetic correlation between the two traits). In the estimation of these parameters, it is convenient to work in terms of the vector $\beta' = 0.5\ [h_1^2\ r_A h_1 h_2\ h_2^2[$ rather than the (2×2) symmetric matrix of genetic variances and covariances.

There is no loss of generality from standardising the traits, for if the diagonal elements of the P matrix are not equal to one, then the phenotypic variables can be standardised using a transformation, T*, with the result that the genetic variance-covariance matrix of the transformed variables is $T^* G T^{*\prime}$. The invariance argument for D-optimality shows that a D-optimal design for the parameter β is also D-optimal for the parameters in $T^* G T^{*\prime}$.

Further, we assume that errors in the phenotypic matrix P can be neglected, either because there is adequate previous data or parental data on which to base estimation of P. The emphasis, within this paper, is on comparing estimation procedures and suggesting designs for genetic parameter estimation and so this assumption should have a negligible effect on the conclusions. Certainly, our formulation leads to known results on univariate heritability estimation.

Estimation and design from response to divergent truncation selection

A common method of estimating genetic parameters for two traits, from divergent truncation selection experiments, is to have two selection groups using a different selection index in each group and measure the selection differentials and the correlated responses for the two traits on both selection indices (Falconer 1981). For each of the selection indices, I_m $(m = 1, 2)$, assume a total of M unrelated individuals are measured for both traits and a proportion p with the highest and p with the lowest index values are selected, such that $pM = N$. A total of RM progeny are reared and recorded and with equal family sizes there are $n = R/2p$ progeny per family. Let i and x be the expected selection differential and abscissa on the standardised normal curve corresponding to p and assume equal selection differentials in the two groups. Note that the upper and lower cases of the letter I denote different parameters, however this is standard notation (Falconer 1981).

Initially alternative estimation procedures and designs will be considered for fixed experimental resources. Later, optimisation of the selection proportion, p, family size, n, and the relative proportion of offspring generation measurements to parental generation measurements, R, will be discussed.

It is of interest to consider the possible combinations of selection weights for the two indices. If a selection index $I_m = b_{1m}x_1 + b_{2m}x_2$, then

$$I_m \equiv (b_{1m}/\sqrt{b_{1m}^2 + b_{2m}^2})\,x_1 + (b_{2m}/\sqrt{b_{1m}^2 + b_{2m}^2})\,x_2$$
$$= x_1\cos\theta_m + x_2\sin\theta_m$$

selects the same individuals, where x_j and b_{jm} are the standardised phenotypic values and index weights of the j^{th} trait for the m^{th} index respectively and $\tan\theta_m = b_{2m}/b_{1m}$. Each selection index is characterised by a single parameter θ_m. By symmetry only the values of θ_m in the range 0° to 180° need consideration. Graphically, the line $x_1 \cos\theta_m + x_2 \sin\theta_m = 0$ makes an angle θ_m with the x_1 axis.

The expected genetic response, ΔG_{jm}, in the progeny for trait j due to selection on index m, is given by the product of the regression of the additive genotype of the j^{th} trait on the phenotype of the m^{th} index and the selection differential (SD_m) of the m^{th} index. Thus $\Delta G_{jm} = 0.5\,(b_{jm}\,\sigma_{jj} + b_{km}\,\sigma_{jk})\,SD_m/\mathrm{var}(I_m)$ where σ_{jj} and σ_{jk} are, respectively, the genetic variance for trait j and genetic covariance for traits j and k (j = 1, 2; k = 3 − j) and var (I_m) is the variance of the m^{th} index. The selection differential for the m^{th} index is calculated as the difference in mean index value between the high and low parental lines. The response in each trait can be estimated as the difference between the high and low progeny lines. The index weights b_{jm} are usually determined by biological arguments about the traits or the desired direction of the response (Eisen 1977). The responses of trait j in selection group m can be written in the form of a regression model, regressing responses in the two measured traits on selection differentials of the indices.

$$\begin{bmatrix}\Delta G_{1m}\\ \Delta G_{2m}\end{bmatrix} = \begin{bmatrix} b_{1m}\,SD_m/\mathrm{var}(I_m) & b_{2m}\,SD_m/\mathrm{var}(I_m) & 0 \\ 0 & b_{1m}\,SD_m/\mathrm{var}(I_m) & b_{2m}\,SD_m/\mathrm{var}(I_m)\end{bmatrix}\beta + [e]$$

or $Y = X\beta + e$. The model can also be defined in terms of selection differentials for each measured trait (SD_{jm}),

$$\begin{bmatrix}\Delta G_{1m}\\ \Delta G_{2m}\end{bmatrix} = \frac{1}{(1-r_p^2)}\begin{bmatrix} SD_{1m} - r_p\,SD_{2m} & -r_p\,SD_{1m} + SD_{2m} & 0 \\ & SD_{1m} - r_p\,SD_{2m} & -r_p\,SD_{1m} + SD_{2m}\end{bmatrix}\beta + [e]$$

where r_p is the phenotypic correlation between the two traits. Alternatively, using the expected value of SD_m, the expected value of the design matrix X can be conveniently written using the angles θ_m,

$$\frac{2i}{\sigma_{1m}}\begin{bmatrix}\cos\theta_m & \sin\theta_m & 0\\ 0 & \cos\theta_m & \sin\theta_m\end{bmatrix} \qquad (1)$$

for each index, where σ_{1m} is the standard deviation of the m^{th} index.

The residuals within lines are correlated, due to the family structure of the design, but there is no correlation of residuals between lines. The 4×4 variance-covariance matrix (V) of the residuals is therefore symmetric and block diagonal

$$V = f\begin{bmatrix}F & 0\\ 0 & F\end{bmatrix}. \qquad (2)$$

The matrix F represents the 2×2 variance-covariance of a family mean after regressing on parental values and the factor f relates the variance of the mean genetic response for one index to the variance of a family mean. With response/selection differential estimation, there are N parents in each of the selected high and low lines, therefore f = 2/N.

The structure of F can be derived using the equations of Hill (1971) for the variance of residuals from single-trait selection. The variance for one progeny mean is

$$F = [(r_{oo}\,G - r_{op}\,G\,P^{-1}\,G\,r_{op}) + (P - r_{oo}\,G)/n] \qquad (3)$$

where r_{oo} and r_{op} Wright's coefficients of relationship for progeny of the same parent and for progeny with parent respectively. Note that the first term in equation (3) is the variance of a family genotypic mean about the regression (drift variance) and the second term is the variance of measurement error in the family mean value. For example, in single-trait selection on parents of one sex with half-sib families

$$G = h^2,\quad P = 1.0,\quad r_{oo} = 0.25,\quad r_{op} = 0.5$$

and

$$F = [0.25\,h^2\,(1 - h^2) + (1 - 0.25\,h^2)/n].$$

Gunsett et al. (1982, 1984) gave similar formulae for V, however their genetic drift term does not include any genetic relationship parameters (r_{oo}, r_{op}) and their measurement error term does not have the divisor of the number of parents in each index.

Investigation of DET (β) and calculation of the inverse of V would be simpler if the matrix F was diagonal. As the matrix F is a function of the genetic and phenotypic variance-covariance matrices, transformation to independent traits would diagonalise F. Such a transformation exists and is often called a canonical transformation (Rao 1973). Let S* be the transformation matrix from the original scale to the canonical scale, such that

$$\begin{bmatrix}C_1\\ C_2\end{bmatrix} = S^*\begin{bmatrix}x_1\\ x_2\end{bmatrix}$$

where C_1 and C_2 are the canonical traits which are phenotypically and genetically uncorrelated. Then S* is such that $S^*PS^{*\prime}$ equals the identity matrix and $S^*GS^{*\prime} = G_C$ where G_C is the diagonal genetic variance-covariance matrix on the canonical scale. For half-sib family data, matrix $F^{-1} = D$ becomes

$$D = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix} \tag{4}$$

where $d_j = (0.25\lambda_j(1-\lambda_j) + (1-0.25\lambda_j)/n)^{-1}$ and λ_j denotes the canonical heritability of the j^{th} canonical trait.

If β_C is the vector of genetic parameters on the canonical scale, similar to β, and the indices on the canonical scale are $I_1 = C_1\cos\theta_{C1} + C_2\sin\theta_{C1}$ and $I_2 = C_1\cos\theta_{C2} + C_2\sin\theta_{C2}$, where θ_{C1} and θ_{C2} are the angles of the canonical selection indices, then $\sigma_{1m} = 1$ and $X'V^{-1}X_C$, the value of $X'V^{-1}X$ for canonical traits, is derived from equations 1 to 4

$$X'V^{-1}X_C = \begin{bmatrix} d_1B_1 & d_1B_3 & 0 \\ d_1B_3 & d_1B_2 + d_2B_1 & d_2B_3 \\ 0 & d_2B_3 & d_2B_2 \end{bmatrix}. \tag{5}$$

The expected value of DET (β_C) is

$$\begin{aligned} \text{DET}(\beta_C) &= |X'V^{-1}X_C| \\ &= d_1d_2(d_1B_2 + d_2B_1)(B_1B_2 - B_3^2) \end{aligned} \tag{6}$$

where

$$\begin{aligned} B_1 &= 2N i^2(\cos^2\theta_{C1} + \cos^2\theta_{C2}) \\ B_2 &= 2N i^2(\sin^2\theta_{C1} + \sin^2\theta_{C2}) = 2N i^2(2 - B_1) \\ B_3 &= N i^2(\sin 2\theta_{C1} + \sin 2\theta_{C2}). \end{aligned} \tag{7}$$

It can be shown that DET $(\beta) = (1-r_p^2)^{-3}$ DET (β_C) (see Appendix 1). In order to maximise DET (β_C), it is differentiated with respect to B_2 and B_3, and the maximum occurs when

$$\frac{B_2}{2N i^2} = \frac{-2(d_1 - 2d_2) \pm 2\sqrt{d_1^2 - d_1d_2 + d_2^2}}{3(d_2 - d_1)} \tag{8}$$

and $B_3 = \sin 2\theta_{C1} + \sin 2\theta_{C2} = 0$, therefore $\theta_{C2} = \theta_{C1} + 90°$ or $\theta_{C1} + \theta_{C2} = 180°$.

There are two cases to consider when maximising DET (β_C). If the canonical heritabilities are equal, d_1 equals d_2, then the maximum value of DET (β_C) occurs when $B_2/2N i^2 = 1$ or $\theta_{C2} = \theta_{C1} + 90°$. The indices on the canonical scale are $I_1 = C_1\cos\theta_{C1} + C_2\sin\theta_{C1}$ and $I_2 = C_2\cos\theta_{C1} - C_1\sin\theta_{C1}$ and as the this pair of axes are at right angles we call this an orthogonal design. The phenotypic covariance between the indices is zero. There are an infinite number of pairs of indices resulting in the maximum value of DET (β_C).

If the canonical heritabilities are unequal, then $\theta_{C1} + \theta_{C2} = 180°$ and θ_{C1} can be derived using equation (8) as $B_2/2N i^2 = 2\sin^2\theta_{C1}$. The indices on the canonical scale are $I_1 = C_1\cos\theta_{C1} + C_2\sin\theta_{C1}$ and $I_2 = C_2\sin\theta_{C1} - C_1\cos\theta_{C1}$. The lines $I_1 = 0$ and $I_2 = 0$ are symmetric about the C_1 and C_2 axes and we call this a symmetric design. Note that the angle between I_1 and the C_1 axis is equal to the angle between I_2 and the C_1 axis.

Estimation and design from offspring-parent regression

The heritability of a trait can be estimated from the regression of progeny performance on parent performance, rather than using a summary of parental information and responses to selection. The design of experiments to estimate the heritability of a trait using offspring-parent regression have been discussed by Hill (1970) and Hill and Thompson (1977).

Offspring-parent regression techniques can be used to estimate genetic parameters of more than one trait simultaneously. The standardised observations on two traits for the j^{th} parent and the mean of its offspring are defined as x_{1j}, x_{2j} and $\bar{o}_{x1j}$, $\bar{o}_{x2j}$, respectively. Then

$$\begin{aligned} \begin{bmatrix} \bar{o}_{x1j} \\ \bar{o}_{x2j} \end{bmatrix} &= 0.5\,G\,P^{-1}\begin{bmatrix} x_{1j} \\ x_{2j} \end{bmatrix} + [e] \\ &= 0.5\,G\begin{bmatrix} s_{1j} \\ s_{2j} \end{bmatrix} + [e] \\ &= \begin{bmatrix} s_{1j} & s_{2j} & 0 \\ 0 & s_{1j} & s_{2j} \end{bmatrix}\beta + [e] \end{aligned}$$

where s_{1j} and s_{2j} are $(x_{1j} - r_p x_{2j})/(1-r_p^2)$ and $(x_{2j} - r_p x_{1j})/(1-r_p^2)$, respectively.

Combining the information from all 4N offspring-parents pairs, β can be estimated as before. The matrix V is now a $8N \times 8N$ block diagonal matrix with the F matrix repeated 4N times down the diagonal.

The contribution of each family to $X'V^{-1}X$ can be expanded as

$$(X'V^{-1}X)_j = s_{1j}^2\begin{bmatrix} D_{11} & D_{12} & 0 \\ D_{21} & D_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix} + s_{1j}s_{2j}\begin{bmatrix} 0 & D_{11} & D_{12} \\ D_{11} & D_{12} + D_{21} & D_{22} \\ D_{21} & D_{22} & 0 \end{bmatrix} + s_{2j}^2\begin{bmatrix} 0 & 0 & 0 \\ 0 & D_{11} & D_{12} \\ 0 & D_{21} & D_{22} \end{bmatrix} \tag{9}$$

where D_{jk} are the elements of F^{-1} (equation (3)). The sums of squares and crossproducts of the parental traits, after selection, are calculated using

$$\begin{aligned}\text{cov}(x_1, x_2 \mid \text{selection on I}) &= \text{cov}(x_1, x_2) - \frac{\text{cov}(x_1, I)\,\text{cov}(x_2, I)}{\text{var}(I)} \\ &\quad + \frac{\text{cov}(x_1, I)\,\text{var}(I^*)\,\text{cov}(x_2, I)}{\text{var}^2(I)} \\ &= \text{cov}(x_1, x_2) + \frac{i\,x\,\text{cov}(x_1, I)\,\text{cov}(x_2, I)}{\text{var}(I)} \qquad (10)\end{aligned}$$

where var (I*) is the variance of the index after selection. The sums of squares and crossproducts of s_{1j} and s_{2j}, after selection, can be determined from

$$\begin{bmatrix} \sum_{j=1}^{4N} s_{1j}^2 & \sum_{j=1}^{4N} s_{1j}s_{2j} \\ \sum_{j=1}^{4N} s_{1j}s_{2j} & \sum_{j=1}^{4N} s_{2j}^2 \end{bmatrix} = P^{-1} \begin{bmatrix} \sum_{j=1}^{4N} x_{1j}^2 & \sum_{j=1}^{4N} x_{1j}x_{2j} \\ \sum_{j=1}^{4N} x_{1j}x_{2j} & \sum_{j=1}^{4N} x_{2j}^2 \end{bmatrix} P^{-1}. \qquad (11)$$

As before, transformation on to the canonical scale results in the diagonalisation of the F matrix and $X'V^{-1}X_C$ has the same structure as in equation (5), where now

$$B_1 = \sum_{j=1}^{4N} C_{1j}^2, \quad B_2 = \sum_{j=1}^{4N} C_{2j}^2 \quad \text{and} \quad B_3 = \sum_{j=1}^{4N} C_{1j}C_{2j} \qquad (12)$$

with C_{1j} and C_{2j} being the observations on the canonical scale of the j^{th} selected parent and a total of 4N selected individuals as before. DET (β_C) becomes

$$\text{DET}(\beta_C) = d_1 d_2 (d_1 B_2 + d_2 B_1)(B_1 B_2 - B_3^2)$$

which is of the same form as equation (6), with B_1, B_2 and B_3 given by equation (12) rather than equation (7). The expected sums of squares and crossproducts of the observations can be rewritten as

$$B_1 = 2N[2 + i\,x(\cos^2\theta_{C1} + \cos^2\theta_{C2})]$$
$$B_2 = 2N[2 + i\,x(\sin^2\theta_{C1} + \sin^2\theta_{C2})]$$
$$B_3 = N\,i\,x[\sin 2\theta_{C1} + \sin 2\theta_{C2}].$$

Then

$$\begin{aligned}\text{DET}(\beta_C) = (2N)^3 d_1 d_2 [&d_1(2 + i\,x\,H) \\ &+ d_2(2 + i\,x(2 - H))][2 + i\,x\,H][2 + i\,x(2 - H)]\end{aligned} \qquad (13)$$

with $H = \sin^2\theta_{C1} + \sin^2\theta_{C2}$. In order to maximise DET (β_C), it is differentiated with respect to B_2 and B_3, and the maximum occurs when

$$H = \frac{-2[(d_1 - 2d_2)\,i\,x - (d_1 + d_2)] \pm 2(2 + i\,x)\sqrt{d_1^2 - d_1 d_2 + d_2^2}}{3\,i\,x(d_2 - d_1)} \qquad (14)$$

and $B_3 = \sin 2\theta_{C1} + \sin 2\theta_{C2} = 0$, therefore $\theta_{C2} = \theta_{C1} + 90°$ or $\theta_{C1} + \theta_{C2} = 180°$.

If the canonical heritabilities are equal, d_1 equals d_2, then H equals one and DET (β_C) is maximised when $\theta_{C2} = \theta_{C1} + 90°$. This corresponds to a ridge of points where DET (β_C) is of constant value (the previously mentioned orthogonal design). If the canonical heritabilities are not equal a symmetric design with $2\sin^2\theta_{C1} = H$, found from equation (14), is again optimal.

The ratio of values of DET (β_C) from the orthogonal design using the offspring-parent regression and response/selection differential estimation is $((2 + i\,x)/i^2)^3 > 1.0$. For example, when p equals 0.10 and 0.20, the ratio equals $(1.38)^3$ and $(1.62)^3$, respectively. The proportional gain in precision $(2 + i\,x)$ from the offspring-parent regression designs comes from two sources. For example, if $I_1 = C_1$ and $I_2 = C_2$, then $(1 + i\,x)$ is proportional to the sums of squares for C_1 from selection on I_1 compared to i^2 used in response/selection differential estimation. The remainder $((2 + i\,x) - (1 + i\,x))$ is proportional to the sums of squares for C_1 with selection on I_2, which is information not used in response/selection differential estimation.

Canonical traits have been used to simplify the development of the variance formulae and interpretation of the designs. When the experiment is being designed, G and hence the canonical transformation are not known precisely, therefore the specification of the optimal design is difficult. However, the class of orthogonal designs includes all pairs of indices that are phenotypically uncorrelated. On the standardised scale, an index $I_2 = x_1 \cos\theta_{C2} + x_2 \sin\theta_{C2}$ can be found phenotypically uncorrelated to $I_1 = x_1 \cos\theta_{C1} + x_2 \sin\theta_{C1}$, if $\tan\theta_{C2} = -(1 + r_p \tan\theta_{C1})/(r_p + \tan\theta_{C1})$. This gives some flexability in the choice of designs. For example, the three pairs of indices $I_1 = x_1$ and $I_2 = x_2 - r_p x_1$, $I_1 = x_2$ and $I_2 = x_1 - r_p x_2$ and also $I_1 = x_1 + x_2$ and $I_2 = x_1 - x_2$ are members of the class of orthogonal designs. This choice of indices can be made without "a priori" knowledge of G and is optimal if the canonical heritabilities are equal.

Manipulation of equations (13) and (14) shows that the ratio of DET (β_C) using the optimal symmetric design compared with using one pair from the above three indices is $(1 + \eta\,\delta)/(1 - \eta^2)$ with

$$\eta = (-1 \pm \sqrt{1 + 3\delta^2})/3 = i\,x(H - 1)/(2 + i\,x)$$

and $\delta = (d_1 - d_2)/(d_1 + d_2)$. For a range of canonical heritabilities, the ratio was generally less than 1.05.

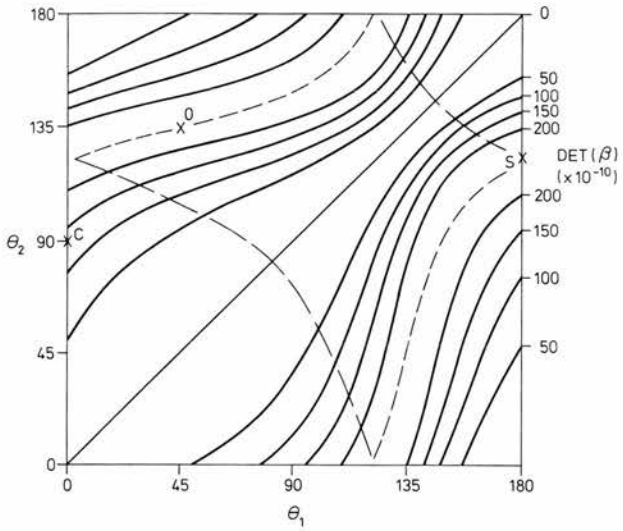


Fig. 1. Contours for DET (β) (divided by 10^{10}) for various linear indices of the traits defined by angles θ_1 and θ_2 with β estimated by response/selection differential. The classes of orthogonal designs (– – –), symmetric designs (——), the orthogonal design $I_1 = x_1 + x_2$, $I_2 = x_1 - x_2$ (O), the optimal symmetric design (S) and the classical design $I_1 = x_1$, $I_2 = x_2$ (C) are included

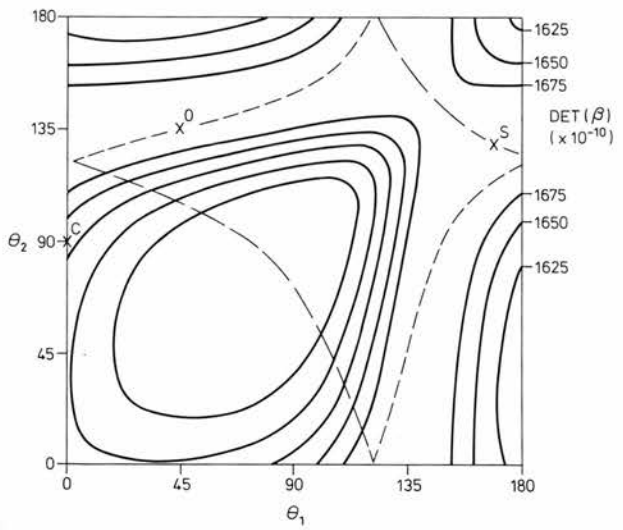


Fig. 2. Contours for DET (β) (divided by 10^{10}) for various linear indices of the traits defined by angles θ_1 and θ_2 with β estimated by offspring-parent regression. Classes of designs and individual designs are denoted as in Fig. 1

This suggests that the proportional increase in precision of a genetic variance or covariance estimate will be at most 0.02 ($= (1.05)^{1/3} - 1$), from using the symmetric design compared with using the orthogonal design. Therefore efficient selection indices can be constructed without estimates of the genetic parameters being available.

To illustrate these results, Figs. 1 and 2 show the contours for DET (β) estimated by response/selection differential (Fig. 1) and offspring-parent regression (Fig. 2). The heritabilities are 0.6 and 0.9 and the genetic and phenotypic correlations are 0.8 and 0.6, respectively, with 600 sires selected per index and a family size of 10 and selection proportion of 0.3, as used by Gunsett et al. (1984). Included are lines indicating the orthogonal designs with the same value of DET (β) (the orthogonal design $I_1 = x_1 + x_2$, $I_2 = x_1 - x_2$ is denoted by 0) and the symmetric designs (the optimal symmetric design is denoted by S). The classical design $I_1 = x_1$, $I_2 = x_2$ is denoted by C ($\theta_1 = 0°$, $\theta_2 = 90°$). The contour for the orthogonal designs in Fig. 1 corresponds to the ridge noted by Gunsett et al. (1984). When $\theta_1 = \theta_2$ in Fig. 1, then DET (β) = 0 because only two parameters can be estimated. The orthogonal, symmetric and classical designs have values of DET (β) (divided by 10^{10}) of 233, 235 and 136 in Fig. 1 and 1,702, 1,719 and 1,635 in Fig. 2, respectively. The ratio $1{,}702/233 = ((2 + i\,x)/i^2)^3 = 1.94^3$ shows the advantage of using offspring-parent regression with orthogonal designs.

When two linear indices are used to select parents, we have shown how to improve the precision of parameter estimates using offspring-parent regression. We have also shown how to choose the linear indices in an efficient way (viz. pairs of orthogonal indices on the canonical scale). We now consider an alternative selection criteria on which to select individuals.

Elliptical selection experimental design

When using offspring-parent regression to estimate genetic parameters, the variance of the genetic parameters depends on the sum of squares of the observations on the parents. When only one trait is of interest, the sum of squares is maximised by selecting individuals with high and low values of the trait to be parents (i.e. selection of individuals with extreme values). By analogy, in the two dimensional case, this suggests selecting a proportion p_E (if the same experimental resources are used as in the divergent selection schemes, then p_E equals 2p) of the 2M individuals measured which are as far from the origin is possible. Invariance arguments suggest using a quadratic index of the form $(x_{1j}\, x_{2j})'\, P^{-1}\, (x_{1j}\, x_{2j})$ for the j^{th} individual. Geometrically, this can be thought of as selecting individuals outside an ellipse given by the formula $(x_1 + x_2)^2/2(1 + r_p) + (x_1 - x_2)^2/2(1 - r_p) = w^2$, where w is chosen such that a proportion p_E of the individuals are outside the ellipse and, because this depends on P, we call the ellipse a phenotypic selection ellipse.

Tallis (1963) considered this type of selection in a different context and showed that the proportion p_E and the variance-covariance matrix of the observations after elliptical selection, P*, can be derived as $p_E = F_2(w^2)$ and $P^* = [F_4(w^2)/F_2(w^2)]\,P$ where $F_k(w^2)$ is the probability that a χ^2 variable with k d.f. is greater than w^2. The recursive procedure of Hill and Pike (1966) gives the relationship between $F_2(w^2)$ and $F_4(w^2)$, viz. $F_4(w^2) = F_2(w^2) + (w^2/2)\exp(-w^2/2)$, where $F_2(w^2) = \exp(-w^2/2) = p_E$. Therefore $P^* = (1 - \log p_E)\,P$.

As before, transformation onto the canonical scale results in the diagonalisation of the F matrix and $\mathrm{DET}(\beta_C)$ can be written as

$$\mathrm{DET}(\beta_C) = (2\,M\,p_E)^3\,d_1\,d_2\,(d_1 + d_2)\,(1 - \log p_E)^3. \quad (15)$$

The ratio of the determinants from elliptical selection and the orthogonal index design is $(2\,(1 - \log p_E)/(2 + i\,x))^3 > 1.0$. For example, when p_E equals 0.2 and 0.4, the ratio equals $(1.23)^3$ and $(1.21)^3$, respectively, which shows the advantage of using the phenotypic selection ellipse rather than selecting on orthogonal canonical indices. Obviously, if no phenotypic selection is performed then $p_E = 1.0$ and $p = 0.5$ and the ratio of the two determinants is one.

The selection criteria $(x_1\,x_2)'\,P^{-1}\,(x_1\,x_2) = w^2$ can be thought as $(x_1 + x_2)^2/2(1 + r_p) + (x_1 - x_2)^2/2(1 - r_p)^2 = w^2$ and $x_1 + x_2$, $x_1 - x_2$ are the axes of the ellipse. For canonical traits the selection ellipse reduces to a canonical circle which is generated by the orthogonal axes $C_1\cos\theta_{C1} + C_2\sin\theta_{C1}$ and $C_2\cos\theta_{C1} - C_1\sin\theta_{C1}$, for all values of θ_{C1}. These axes are precisely those of the orthogonal indices suggested in the previous section. This naturally leads to the question if a canonical ellipse generated by the symmetric axes $C_1\cos\theta_{C1} + C_2\sin\theta_{C1}$ and $C_1\sin\theta_{C1} - C_2\cos\theta_{C1}$ can give a more efficient design. The calculation of the sums of squares and crossproducts for the parental values is more difficult and requires numerical integration (see Appendix 2 for calculation of $\mathrm{DET}(\beta_C)$). The maximum value of $\mathrm{DET}(\beta_C)$ occured when the canonical ellipse was rotated by an angle ϕ with values 0° and 90°, for $0° \le \phi \le 180°$. When $\phi = 90°$, this correspounds to reparameterising C_1 as C_2 and vica versa. Therefore, the canonical ellipse generated by the symmetric axes gives the most efficient design.

Again there is the difficulty that these axes require estimates of G and we could not find an analytic formula for the optimal angle. The ratio of values of $\mathrm{DET}(\beta_C)$ from using the optimal symmetric and orthogonal axes depends on the proportion of individuals selected as parents. For combinations of canonical heritabilities in the range of 0.1 to 0.9 and a range of selection proportions ($0.05 < p_E < 0.30$), the maximum value of the ratio was 1.01. The ratio decreased as the selection proportion increased and as the magnitude of the difference between $\lambda_1(1 - \lambda_1)$ and $\lambda_2(1 - \lambda_2)$ decreased. Therefore, there is a negligible loss of efficiency when using the phenotypic elliptical selection scheme compared with using the optimal elliptical scheme.

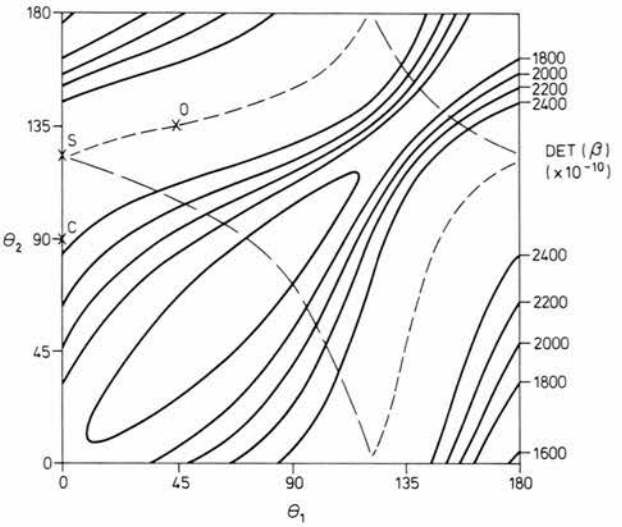


Fig. 3. Contours for $\mathrm{DET}(\beta)$ (divided by 10^{10}) for various quadratic indices of the traits defined by angles θ_1 and θ_2 with β estimated by offspring-parent regression. Classes of designs and individual designs are denoted as in Fig. 1

Figure 3 shows $\mathrm{DET}(\beta)$ using ellipses generated by axes $I_1 = x_1\cos\theta_1 + x_2\sin\theta_1$ and $I_2 = x_1\cos\theta_2 + x_2\sin\theta_2$ using the same G and P matrices and experimental facilities as in Figs. 1 and 2. The values (divided by 10^{10}) of $\mathrm{DET}(\beta)$ for the orthogonal (O), symmetric (S) and classical (C) axes are 2,650, 2,652 and 2,454, respectively, showing a marked increase over the corresponding values in Fig. 2, with

$$2{,}650/1{,}702 = (2\,(1 - \log p_E)/(2 + i\,x))^3 = (1.16)^3,$$

for the phenotypic selection ellipse.

Optimising the selection proportion, the family size and the ratio of individuals measured in the two generations

If the canonical heritabilities are equal, say to λ, the optimum proportion to select for maximising $\mathrm{DET}(\beta)$ with different estimation methods and selection designs can be found. For example, if β is estimated by response/selection differential the optimal p is found by differentiation of equation (6) with respect to p. The solution is given by

$$\frac{(1 - r_{oo}\lambda)}{R\,\lambda\,(r_{oo} - r_{op}^2\lambda)} = \frac{2x - i}{4\,(i - x)\,p}$$

which suggests that p must be at least 0.27, that is when $2x > i$.

When estimating genetic parameters using offspring-parent regression, the optimal proportion p is obtained by differentiating equation 12 with respect to p, which satisfies

$$\frac{(1-r_{oo}\lambda)}{R\lambda(r_{oo}-r_{op}^2\lambda)}=\frac{1+x^2}{2p(1+ix-x^2)}=W(2,p) \qquad (16)$$

which is similar to that of Hill and Thompson (1977), derived in a univariate context,

$$\frac{(1-r_{oo}\lambda)}{R\lambda(r_{oo}-r_{op}^2\lambda)}=\frac{x^2}{2p(1+ix-x^2)}=W(1,p)\,.$$

The minimum value of the right hand side of $W(2,p)$ is one when $p=0.5$, and all individuals are then used as parents. When using a phenotypic selection ellipse, differentiating equation (15) with respect to p_E, gives the result

$$\frac{(1-r_{oo}\lambda)}{R\lambda(r_{oo}-r_{op}^2\lambda)}=\frac{-\log p_E}{p_E}=W(3,p_E)\,. \qquad (17)$$

These equations give an optimal design for fixed numbers of individuals in the parental, 2M, and offspring, 2MR, generations. If the balance of individuals in the two generations can be adjusted, R, then the optimal value of $DET(\beta_C)/(2M(1+R))^3$, a measure of the efficiency of the design on a per individual measured basis, can be determined. When divergent selection lines are used, the optimum value of p satisfies

$$\frac{(1-r_{oo}\lambda)}{\lambda(r_{oo}-r_{op}^2\lambda)}=\left[\frac{1+x^2}{1+ix-x^2}\right]^2\frac{1}{2p}=W(4,p)$$

and $R=(1+x^2)/(1+ix-x^2)$. When the phenotypic selection ellipse is used, the optimum value of p_E satisfies

$$\frac{(1-r_{oo}\lambda)}{\lambda(r_{oo}-r_{op}^2\lambda)}=(\log p_E)^2/p_E=W(5,p_E) \qquad (18)$$

and $R=-\log p_E$. Figure 4 has been constructed to aid in the solution of the above equations, giving values of $W(s,q)$ against the total proportion selected, p_T, where $q=p_{T/2}$ for $s=1,2$ and 4 and $q=p_T$ for $s=3$ and 5.

Since the genetic parameters are not known "a priori", designs should be robust to poor estimates of these parameters. The $DET(\beta_C)$ values using elliptical selection were calculated for a range of equal canonical heritabilities, with fixed values of R, at fixed and optimum values of p_E and were then compared with $DET(\beta_C)$ values when both the p_E and R are optimised (Fig. 5). The efficiency of designs when both p_E and R are optimised are shown as 100 and $DET(\beta_C)$ values of other designs are shown relative to this base. Figure 5 indicates that for a wide range of canonical heritabilities, with $R=2$, $p_E=0.20$ is efficient. For example, with λ values in the ranges (0.18, 0.87) and (0.13, 0.90) designs using $p_E=0.20$ are at least 0.90 and 0.95 as efficient as the optimal design. When $R=10$, designs are generally less than 0.40 as efficient as when R is optimised, although $p_E=0.30$ is close to the optimal value of p, for $R=10$.

The optimum proportion of individuals to select as parents has been determined, but only when the

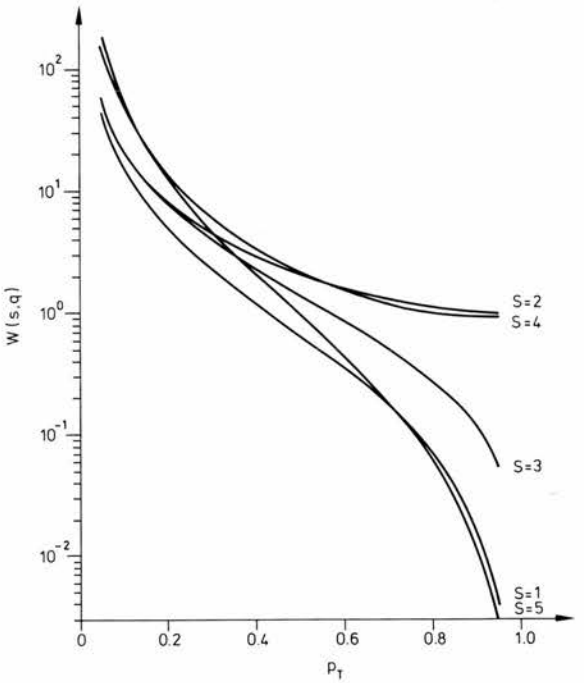


Fig. 4. Values of $W(s,q)$ plotted against the total proportion selected, p_T

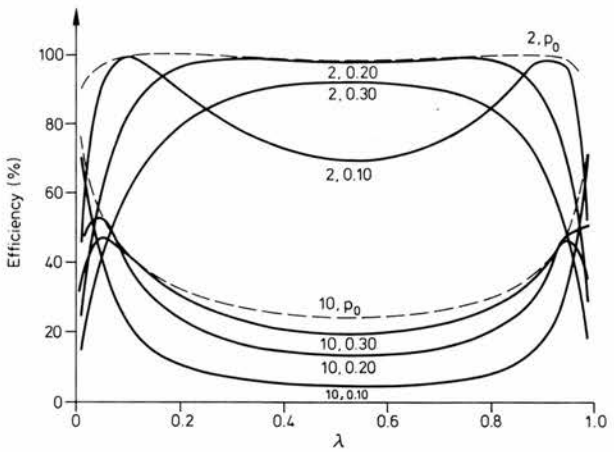


Fig. 5. Efficiency of alternative elliptical designs expressed as $DET(\beta)$ for a fixed total number recorded relative to that when both p_E and R are optimised. Results are given for specified R and p_E, with p_E fixed or optimal (p_O)

canonical heritabilities are equal. When the canonical heritabilities are unequal, one suggestion is to use a pooled value of λ in equations (16) to (18), with λ chosen such that the resulting d satisfies

$$2\,d^3 = d_1\,d_2\,(d_1 + d_2)\,. \qquad (19)$$

As there are two solutions to the quadratic equation for λ, we suggest using the λ value that lies between λ_1 and λ_2. Due to some symmetry in the d value (i.e. $\lambda\,(1-\lambda)$), the value of λ is less than 0.5 when $\lambda_1+\lambda_2<1$ and λ is greater than 0.5 otherwise. The value of λ satisfying equation (19) is essentially independent of the value of n, the number of progeny per parent, when n is moderate (>15). When no "a priori" estimates of the genetic parameters are available, $n=25$ seems a reasonable value to estimate λ with. The values of $\mathrm{DET}(\beta)$ calculated with the optimum p_E were regressed on the $\mathrm{DET}(\beta)$ values calculated using p_E derived from equation (17), for combinations of canonical heritabilities in the range 0.1 to 0.9 with various R and n values. If the methods of choosing p_E were identical, then the pooled regression coefficient and intercept are expected to have values 1.0 and 0.0 and the actual values were 0.980 and 0.003, respectively. Therefore, the use of equation (19) to generate a pooled λ value seems reasonable, for estimation of the optimum selection proportion, p_E.

Extensions

The gains from using assortative mating when selection is practiced on both male and female parents in one dimensional problems have been demonstrated (Reeve 1955; Hill and Thompson 1977). The same results apply directly to multivariate designs with selection of mates being based on minimising the "phenotypic distance" between mates.

Selection over several generations can also be effective in increasing the precision. However, the distribution of the progeny measurements, the next parental generation, would not be normal which introduces further complications in the estimation of the variance of the parameters.

Estimation of genetic parameters with a selection ellipsoid is not just limited to two traits. For $v\,(>2)$ traits the phenotypic selection ellipsoid and transformation onto the canonical scale can be used as before. When the traits have equal canonical heritabilities, the determinant of the inverse of the variance-covariance matrix of the genetic parameter estimates, on the canonical scale, can be written as

$$\mathrm{DET}(\beta_C) = (v\,M\,p_E\,d\,[F_{v+2}(w^2)/F_v(w^2)])^{v(v+1)/2}\,2^{v(v-1)/2}\,.$$

The optimum proportion of individuals to select can be determined by differentiating $v\,M\,p_E\,d\,F_{v+2}(w^2)/F_v(w^2) = v\,M\,p_E\,d\,K$ with respect to p_E in order to maximise the value of $\mathrm{DET}(\beta_C)$, where $K=F_{v+2}(w^2)/F_v(w^2)$. However by defining the function $W(p_E)$, the optimal proportion is determined by solving

$$\frac{(1-r_{oo}\lambda)}{R\,\lambda\,(r_{oo}-r_{op}^2\lambda)} = \frac{-1}{p_E}\left[\frac{K}{p_E\,\partial K/\partial p_E}+1\right] = W(p_E)$$

where $v\,M\,p_E$ is the total number of individuals selected for the ellipsoidal design. The mean parental sums of squares decreases as the number of traits increases and obviously as the proportion selected decreases. However marked gains for increasing the precision of estimates of genetic parameters can be made with at least 5 traits.

An example

An example of a design using elliptical selection is taken from an ABRO sheep experiment to estimate genetic parameters for growth rate and carcass leanness in lambs slaughtered at fixed age. A total of 100 rams are measured and 750 progeny are expected, giving a R value of 7.5. The "a priori" estimates of the heritabilities are 0.20 and 0.40 and the genetic and phenotypic correlations are 0.25 and 0.15, respectively. The canonical traits are $1.010\,x_1 - 0.203\,x_2$ and $0.052\,x_1 + 0.991\,x_2$, which are phenotypically uncorrelated and have phenotypic variance of 1.0. The canonical heritabilities are 0.192 (derived from $1.010^2 h_1 + 2\,(1.010)\,(-0.203)\,r_A\,h_1\,h_2 + (-0.203)^2 h_2^2$) and 0.401, and using $n=25$ to estimate λ, the value of 0.262 is derived from equation (19). Given the R value of 7.5, the optimum proportion of rams to select, p_E, is 0.378 from solving $W(3, p_E) = (-\log p_E)/p_E = 2.57$ (equation (17)) or using Fig. 4, and so each selected ram has an expected 20 progeny. Therefore 38 rams are selected such that $x_1^2 + 2\,(-0.15)\,x_1\,x_2 + x_2^2 > 1.94\,(1-0.15^2)$ where x_j are the standardised measurements of growth rate and carcass leanness. The value of $w^2 = 1.94$ is derived from $p_E = \exp(-w^2/2)$.

The matrix $X'V^{-1}X_C$ on the canonical scale can be derived using equations (5) and (12) and is diagonal with elements 866, 1,578 and 712 using $d_1 = 11.58$ and $d_2 = 9.52$ with $B_1 = B_2 = 38\,(1-\log 0.38)$ and $B_3 = 0$. Appendix 1 derives the matrix R* such $R^*\beta = \beta_C$, and in this case

$$R^* = \begin{bmatrix} 0.102 & -0.411 & 0.041 \\ 0.053 & 0.990 & -0.202 \\ -0.003 & 0.104 & 0.982 \end{bmatrix}.$$

The variance-covariance matrix of the genetic parameter estimates is then

$$4\,(R^*)^{-1}\,(X'V^{-1}X_C)^{-1}\,((R^*)^{-1})' = 4\,\mathrm{var}(\beta)$$

$$= \begin{bmatrix} 46.4 & 7.8 & 1.3 \\ 7.8 & 26.0 & 8.7 \\ 1.3 & 8.7 & 56.0 \end{bmatrix} \times 10^{-4}\,.$$

The expected standard errors for the heritabilities of 0.20 and 0.40 are 0.068 and 0.075, respectively and for the genetic covariance of 0.064 the standard error is 0.051.

If the rams were split into two groups and selected high and low in each group, using an orthogonal design, then the

variances of the genetic parameter estimates are proportionately increased by 1.21 (derived from $2(1-\log p_E)/(2+i\,x)$) compared to using elliptical selection. If only information on the parental selection traits is used, then the proportional increase is larger, 1.75 from $2(1-\log p_E)/(1+i\,x)$.

If the classical design is used to estimate the genetic parameters, then the matrix $X'V^{-1}X$, determined from equations (9) and (10), equals

$$\begin{bmatrix} 743 & -187 & 11 \\ -187 & 1{,}381 & -175 \\ 11 & -175 & 616 \end{bmatrix}$$

using $D_{11}=11.85$, $D_{22}=9.83$ and $D_{12}=D_{21}=-1.88$ with

$$\sum_{j=1}^{38} x_{1j}^2 = \sum_{j=1}^{38} x_{2j}^2 = 19\,(1+i\,x) + 19\,(1+i\,x\,(0.15)^2) = 62.36$$

and

$$\sum_{j=1}^{38} x_{1j}\,x_{2j} = 12.84\,.$$

Then

$$\sum_{j=1}^{38} s_{1j}^2 = \sum_{j=1}^{38} s_{2j}^2 = 62.7 \quad \text{and} \quad \sum_{j=1}^{38} s_{1j}\,s_{2j} = -5.83$$

from equation (11). The variance-covariance matrix of the genetic parameters is then 4 var (β), as before, and equals

$$\begin{bmatrix} 55.8 & 7.7 & 1.2 \\ 7.7 & 31.1 & 8.7 \\ 1.2 & 8.7 & 67.4 \end{bmatrix} \times 10^{-4}.$$

Therefore, the proportional increase in the variance of the genetic parameter estimates using the classical design compared to the elliptical design is $1\cdot 22$.

Note that the matrix of weights on the original scale contributing to the selection indices (B) can be determined from the matrix of weights on the canonical scale (ANG_C). If selection is on the orthogonal canonical indices $I_1 = C_1 + C_2$ and $I_2 = C_1 - C_2$, such that $\theta_{C1}=45°$ and $\theta_{C2}=135°$, then

$$ANG_C = \begin{bmatrix} \cos\theta_{C1} & \sin\theta_{C1} \\ \cos\theta_{C2} & \sin\theta_{C2} \end{bmatrix} = \begin{bmatrix} 0.707 & 0.707 \\ -0.707 & 0.707 \end{bmatrix}$$

and

$$B = ANG_C\,S^* = \begin{bmatrix} 0.751 & 0.557 \\ -0.677 & 0.844 \end{bmatrix}$$

Equivalent indices are

$$\begin{bmatrix} \cos\theta_1 & \sin\theta_1 \\ \cos\theta_2 & \sin\theta_2 \end{bmatrix} = \begin{bmatrix} 0.803 & 0.596 \\ -0.626 & 0.780 \end{bmatrix}$$

and the angles of the indices on the original scale are 36.6° and 128.8°.

References

Eisen EJ (1977) Antagonistic selection index results with mice. In: Pollak E, Kempthorne O, Bailey TB (eds) Proc Int Conf Quant Genet. Iowa State University, Ames, Iowa, USA, pp 117–139

Falconer DS (1981) Introduction to quantitative genetics, 2nd edn. Longman, London

Gunsett FC, Andriano KN, Rutledge JJ (1982) Estimating the precision of estimates of genetic parameters realised from multiple-trait experiments. Biometrics 38:981–989

Gunsett FC, Andriano KN, Rutledge JJ (1984) Estimation of genetic parameters from multiple trait selection experiments. J Anim Sci 58:591–599

Hill ID, Pike MD (1967) Chi-squared integral. Algorithm 299. Commun Ass Comput Mach 10:243–244

Hill WG (1970) Design of experiments to estimate heritability by regression of offspring on selected parents. Biometrics 26:566–571

Hill WG (1971) Design and efficiency of selection experiments for estimating genetic parameters. Biometrics 27: 293–312

Hill WG, Thompson R (1977) Design of experiments to estimate offspring-parent regression using selected parents. Anim Prod 24:163–168

Rao CR (1973) Linear statistical inference and its applications. Second edition. Wiley and Sons, New York

Reeve ECR (1955) The variance of the genetic correlation coefficient. Biometrics 11:357–374

Robertson A (1959) Experimental design in the evaluation of genetic parameters. Biometrics 15:219–226

Silvey SD (1980) Optimal designs. Chapman and Hall, London

Tallis GM (1959) Sampling errors of genetic correlation coefficients calculated from analyses of variance and covariance. Aust J Stat 1:35–43

Tallis GM (1963) Elliptical and radial truncation in normal populations. Ann Math Stat 34:940–944

Thompson R (1976) Design of experiments to estimate heritability when observations are available on parents and offspring. Biometrics 32:283–304

Appendix 1

The value of DET (β) can be determined from DET (β_C). Since $G_C = S^* G S^{*\prime}$, then

$$\beta_C = 0.5 \begin{bmatrix} G_{C11} \\ G_{C12} \\ G_{C22} \end{bmatrix} = \begin{bmatrix} S_{11}^{*2} & 2S_{11}^* S_{12}^* & S_{12}^{*2} \\ S_{11}^* S_{21}^* & S_{11}^* S_{22}^* + S_{21}^* S_{12}^* & S_{12}^* S_{22}^* \\ S_{21}^{*2} & 2S_{21}^* S_{22}^* & S_{22}^{*2} \end{bmatrix}.$$

Let the above 3×3 matrix be denoted R^*, then

$\beta_C = R_\beta^*$ and $\beta = (R^*)^{-1}\,\beta_C$

$\text{var}(\beta) = (R^*)^{-1}\,\text{var}(\beta_C)\,((R^*)^{-1})'$

$\text{var}(\beta)^{-1} = R^*\,\text{var}(\beta_C)^{-1}\,R^{*\prime}$

$|X'V^{-1}X| = |R^*|^2\,|X'V^{-1}X|_C\,.$

As $S^* P S^{*\prime} = I$, then $|S^*|^{-2} = |P| = (1-r_p^2)$. The determinant of R^* can be shown to equal $|S^*|^3$, therefore

$$\text{DET}(\beta) = (1-r_p^2)^{-3}\,\text{DET}(\beta_C)\,.$$

Appendix 2

In this appendix the calculation of p_E and the mean sums of squares and crossproducts after elliptical truncation selection is illustrated. The selection ellipse based on symmetric axes is $C_1\cos\theta + C_2\sin\theta$ and $C_1\cos\theta - C_2\sin\theta$ is $a^2C_1^2 + b^2C_2^2 = w^2$, where C_1, C_2 are the canonical variates and $a^2 = 2\cos^2\theta$, $b^2 = 2\sin^2\theta$. Given the proportion to be selected, p_E, the

"size" of the ellipse, w, satisfies

$$p_E = 1 - \frac{4}{\sqrt{2\pi}} \int_0^{\sqrt{a w^2}} \exp(-C_1^2/2) \int_0^{C_2(C_1)} \frac{1}{\sqrt{2\pi}} \cdot \exp(-C_2^2/2)\, dC_2\, dC_1$$

where $C_2(C_1) = \sqrt{(w^2 - a^2 C_1^2)/b^2}$. Likewise the mean sum of squares and crossproducts of the canonical variates after elliptical selection are given by $S S_1(\theta)$, $S S_2(\theta)$ and $C P(\theta)$

$$p_E\, S S_1(\theta_1) = 4 \left[\frac{1}{\sqrt{2\pi}} \int_0^{\sqrt{a w^2}} C_1^2 \exp(-C_1^2/2) \int_{C_2(C_1)}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot \exp(-C_2^2/2)\, dC_2\, dC_1 + \frac{1}{\sqrt{2\pi}} \int_{a w^2}^{\infty} C_1^2 \exp(-C_1^2/2)\, 0.5\, dC_1 \right]$$

with $S S_2(\theta_1) = S S_1(90° - \theta_1)$ and by symmetry $C P(\theta) = 0$. By integrating by parts,

$$p_E\, S S_1(\theta_1) = 4 \left[\frac{1}{2} \int_0^{\sqrt{a w^2}} C_1^2 \exp(-C_1^2/2)\, p_2\, dC_1 + x\, z + p_1 \right] \quad (A1)$$

where

$$p_1 = \frac{1}{2\sqrt{2\pi}} \int_0^{\sqrt{a w^2}} \exp(-C_1^2/2)\, dC_1 \quad \text{and}$$

$$p_2 = \frac{1}{\sqrt{2\pi}} \int_{C_2(C_1)}^{\infty} \exp(-C_2^2/2)\, dC_2$$

and z is the height of the ordinate at truncation point x. If the indices of the selection ellipse are defined by angles θ_{C1} and θ_{C2}, the ellipse can be written as:

$$w^2 = (C_1 \cos\theta_{C1} + C_2 \sin\theta_{C1})^2 + (C_1 \cos\theta_{C2} + C_2 \sin\theta_{C2})^2$$

or

$$w^2 = 2u^2 \cos^2((\theta_{C1} - \theta_{C2})/2) + 2v^2 \sin^2((\theta_{C1} - \theta_{C2})/2)$$

where

$$u = C_1 \cos((\theta_{C1} + \theta_{C2})/2) + C_2 \sin((\theta_{C1} + \theta_{C2})/2)$$

$$v = C_1 \cos((\theta_{C1} + \theta_{C2} + 180°)/2) + C_2 \sin((\theta_{C1} + \theta_{C2} + 180°)/2)$$

which is the equation of an ellipse on a scale with orthogonal axes u and v. The sums of squares of u and v ($S S_u$ and $S S_v$) can therefore be calculated using equation (A1). By transforming back to the canonical scale, the mean sums of squares and crossproducts of the canonical variates are

$$S S_1 = S S_u \cos^2((\theta_{C1} + \theta_{C2})/2) + S S_v \sin^2((\theta_{C1} + \theta_{C2})/2)$$

$$S S_2 = S S_u \sin^2((\theta_{C1} + \theta_{C2})/2) + S S_v \cos^2((\theta_{C1} + \theta_{C2})/2)$$

$$C P = (S S_u - S S_v)(\sin(\theta_{C1} + \theta_{C2}))/2\,.$$

18

The estimation of heritability with unbalanced data. I. Observations available on parents and offspring

by

R. Thompson

BIOMETRIC *33*, 485–495
September, 1977

The Estimation of Heritability with Unbalanced Data I. Observations Available on Parents and Offspring

ROBIN THOMPSON

A.R.C. Unit of Statistics, University of Edinburgh, Kings Buildings Edinburgh EH9 3JZ, Scotland

Summary

The maximum likelihood (ML) estimation of heritability when data are available on both parents and offspring and the offspring data have a hierarchical structure is discussed. Computational forms are developed that are useful when the families of full-sibs and of half-sibs are of unequal sizes. Extensions to other variance component situations are discussed.

1. Introduction

Hill and Nicholas (1974) and Thompson (1976) have recently discussed the design of experiments to estimate heritability when data are available on parents and offspring. They show the advantage of maximum likelihood (ML) methods over regression and sib covariance methods. They consider the efficiency of various balanced hierarchical designs, that is designs in which the same number of dams are mated to each sire and the same number of offspring are measured from each dam. In practice this balance is seldom attained and in this paper we consider estimation of heritability for unbalanced data.

In balanced designs it is useful to divide the data into five parts due to differences within dams, between dams within sires, between sires with offspring, between animals with no offspring and between groups of animals with and without offspring (Thompson 1976). This partition suggests estimators in the unbalanced case but the weighting to be given to each full-sib and half-sib mean has yet to be decided. The weighting of the means has been discussed previously, in the context of using only the parent-offspring covariances, for example by Kempthorne and Tandon (1953) and Ollivier (1974), and in the context of using the sib covariances alone, for example Robertson (1962). The weighting to be given to the means is not obvious when we wish to use both the parent-offspring and sib covariances efficiently and we use instead an ML procedure. We try to take advantage of the special structure of the data, noting that the formulation for the unbalanced case is completely different from the formulation for the balanced case.

In Section 2 we discuss a model for the unbalanced case and give a suitable form for the variance-covariance structure. In Section 3 we use this structure to indicate how the parameters might be estimated. In Section 4 we indicate the modifications necessary when different models are assumed. The method we develop can be useful in the more general problem of estimating variance components and we discuss this in Section 4.

Key Words: Heritability estimation; Maximum likelihood estimation; Unbalanced data.

2. *The Model*

In this section we introduce a linear model in order to derive a convenient form for the variance matrix of the of the observations. As far as possible we follow the notation of Thompson (1976). We assume that we have observations on s' males and D' females, and that the males and females are unrelated. Suppose s of the s' males are mated to females and offspring are raised from each mating, and that D females in all are used as mothers. Let x_i be the measurement on the i *th* male; without loss of generality we can let $1 \leq i \leq s$ represent males that have offspring and $s < i \leq s'$ represent males with no offspring. Let y_{ij} be the measurement of the j *th* dam mated to the i *th* sire, and z_{ijk} be the measurement on the k *th* offspring. Let n_i be the number of dams mated to the i *th* sire and n_{ij} be the number of offspring of the j *th* dam mated to the i *th* sire. Hence the suffixes i, j and k on y and z run from one to s, n_i and n_{ij}, respectively. Let $y_{(s+1)j}$ represent the measurements on the $D' - D$ females that have no recorded offspring ($1 \leq j \leq D' - D$).

We assume that the observations are multivariate normally distributed with mean values given by

$$E(\mathbf{x}) = \mathbf{X}_m\alpha, \qquad E(\mathbf{y}) = \mathbf{X}_f\alpha, \qquad E(\mathbf{z}) = \mathbf{X}_o\alpha, \tag{2.1}$$

where $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are vectors consisting of the measurements x_i, y_{ij} and z_{ijk}, α is a vector of t unknown parameters representing fixed effects such as season, year, etc. and $\mathbf{X}_m$, $\mathbf{X}_f$ and $\mathbf{X}_o$ are known matrices relating the fixed effects to the male, female and offspring data.

We assume that the variance of an observation is σ^2_P independent of sex and generation. We assume that the covariances between parent and offspring, between full-sibs and between half-sibs are $\sigma^2_A/2$, $\sigma^2_A/2 + \sigma^2_K$ and $\sigma^2_A/4$ respectively. σ^2_A, representing the additive genetic variance. σ^2_K represents the covariance between full-sibs other than that due to additive genetic variance and so includes terms due to common environment and dominance. Inter-locus interactions (e.g. epistasis) are assumed not to exist. The maternal environmental covariance between parent and offspring is assumed to be zero, but the argument developed can be extended to include such a term in the model. We assume that there is no covariance between offspring of different sires. Heritability, h^2, which is the ratio of additive genetic variance to the total variance, can be written as σ^2_A/σ^2_P in this parameterization.

Now we group the data according to the sire with which they are associated so that we can write the variance-covariance matrix, $\mathbf{V}$, in a convenient block diagonal form. This form is highly structured and we will later exploit this structure when we solve the estimating equations. The data for the i *th* sire can be written as a $(m_i \times 1)$ vector $\mathbf{w}_i$, where $\mathbf{w}_i$ has elements x_{ij}, y_{ij} $(j = 1, \cdots, n_i)$ and z_{ijk} $(j = 1, \cdots, n_i; k = 1, \cdots, n_{ij})$, and $m_i = 1 + \Sigma_j(1 + n_{ij})$. Therefore $\mathbf{w}_i$ consists of the records of the i *th* sire, of his n_i mates and their n_{ij} offspring. For instance, suppose the first sire is mated to three dams and three, two and one offspring are raised from the three dams. Then in the notation of this paper space $n_1 = 3, n_{11} = 3, n_{12} = 2$ and $n_{13} = 1$, hence

$$m_1 = 1 + 3 + 3 + 2 + 1 = 10$$

and

$$\mathbf{w}'_1 = (x_{11}, y_{11}, y_{12}, y_{13}, z_{111}, z_{112}, z_{113}, z_{121}, z_{122}, z_{131}).$$

When the data are so arranged the covariances between the elements of $\mathbf{w}_i$ and $\mathbf{w}_j$ are zero $(i \neq j)$.

In order to obtain a succinct representation for the variance-covariance matrix for $\mathbf{w}_i$, $\mathbf{V}_i$, we introduce a linear model for $\mathbf{w}_i$ of the form

$$\begin{aligned} x_i &= \mathbf{X}_{mi}\alpha + \beta_{mi} + e_i, \\ y_{ij} &= \mathbf{X}_{fij}\alpha + \beta_{fij} + e_{ij}, \end{aligned} \tag{2.2}$$

where $\mathbf{X}_{mi}$ represents the rows of $\mathbf{X}_m$ associated with the *i th* male. Similarly, $\mathbf{X}_{fij}$ and $\mathbf{X}_{oijk}$ represent the rows of $\mathbf{X}_f$ and $\mathbf{X}_o$ related to the *(ij) th* dam and *(ijk) th* offspring, β_{mi} and β_{fij} represent male and female effects associated with the parental observations and β_{si} and β_{dij} represent sire and dam effects associated with the offspring observations.

We suppose e_i, e_{ij} and e_{ijk} are independently normally distributed with variance σ^2 and that the variance matrices of (β_{mi}, β_{si}) and $(\beta_{fij}, \beta_{dij})$ are $\mathbf{\Sigma}_s\sigma^2$ and $\mathbf{\Sigma}_d\sigma^2$ and that all other covariances are zero. We write $\mathbf{\Sigma}_s\sigma^2$ and $\mathbf{\Sigma}_d\sigma^2$ in the form

$$\mathbf{\Sigma}_s\sigma^2 = \begin{pmatrix} \theta_1 & \theta_2 \\ \theta_2 & \theta_3 \end{pmatrix} \quad \text{and} \quad \mathbf{\Sigma}_d\sigma^2 = \begin{pmatrix} \theta_4 & \theta_5 \\ \theta_5 & \theta_6 \end{pmatrix} \tag{2.3}$$

We now relate these parameters to the genetic parameters by writing the covariances between relatives in terms of both the θ's and the genetic parameters. We require

$$\begin{aligned}
&\text{var}\,(x_i) = \sigma^2{}_P = \text{var}\,(\beta_{mi}) + \text{var}\,(e_i) && = \theta_1 + \sigma^2,\\
&\text{var}\,(y_{ij}) = \sigma^2{}_P = \text{var}\,(\beta_{fij}) + \text{var}\,(e_{ij}) && = \theta_4 + \sigma^2,\\
&\text{var}\,(z_{ijk}) = \sigma^2{}_P = \text{var}\,(\beta_{si}) + \text{var}\,(\beta_{dij}) + \text{var}\,(e_{ijk}) && = \theta_3 + \theta_6 + \sigma^2,\\
&\text{cov (sire, offspring)} = \tfrac{1}{4}\sigma^2{}_A = \text{cov}\,(\beta_{mi}, \beta_{si}) && = \theta_2,\\
&\text{cov (dam, offspring)} = \tfrac{1}{4}\sigma^2{}_A = \text{cov}\,(\beta_{fi}, \beta_{di}) && = \theta_5,\\
&\text{cov (paternal half sibs)} = \tfrac{1}{4}\sigma^2{}_A = \text{var}\,(\beta_{si}) && = \theta_3,\\
&\text{cov (full sibs)} = \tfrac{1}{2}\sigma^2{}_A + \sigma^2{}_K = \text{var}\,(\beta_{si}) + \text{var}\,(\beta_{dij}) && = \theta_3 + \theta_6,
\end{aligned}$$

hence

$$\begin{aligned}
&\sigma^2 = \sigma^2{}_P - \tfrac{1}{2}\sigma^2{}_A - \sigma^2{}_K, \qquad \theta_1 = \theta_4 = \tfrac{1}{2}\sigma^2{}_A + \sigma^2{}_K,\\
&\theta_2 = \theta_5 = \tfrac{1}{2}\sigma^2{}_A, \qquad \theta_3 = \tfrac{1}{4}\sigma^2{}_A, \quad \text{and} \quad \theta_6 = \tfrac{1}{4}\sigma^2{}_A + \sigma^2{}_K.
\end{aligned} \tag{2.4}$$

Therefore, we see that the model (2.2) can be made consistent with the covariance structure derived by genetic arguments. By use of (2.4) in (2.3) we note that $\mathbf{\Sigma}_d\sigma^2$ is not necessarily positive semi-definite.

Equation (2.2) can be written compactly in matrix form as

$$\mathbf{w}_i = \mathbf{X}_i\boldsymbol{\alpha} + \mathbf{Z}_i\boldsymbol{\beta}_i + \mathbf{e}_i \tag{2.5}$$

where $\boldsymbol{\beta}_i$ is a $1 \times 2(1 + n_i)$ vector with first two elements β_{mi}, β_{si} and the next $2n_i$ elements are β_{fij}, β_{dij} $(j = 1, \cdots, n_i)$. Z_i is a matrix with elements zero or one relating the $\boldsymbol{\beta}_i$ effects to $\mathbf{w}_i$. In the previous example

$$\mathbf{Z}_1 = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix} \quad \text{and} \quad \boldsymbol{\beta}_1 = \begin{bmatrix}
\beta_{m1} \\ \beta_{s1} \\ \beta_{f11} \\ \beta_{d11} \\ \beta_{f12} \\ \beta_{d12} \\ \beta_{f13} \\ \beta_{d13}
\end{bmatrix}$$

The variance matrix of $\mathbf{w}_i$, $\mathbf{V}_i$, can now be written as

$$\mathbf{V}_i = (\mathbf{Z}_i\boldsymbol{\Gamma}_i\mathbf{Z}'_i + \mathbf{I})\sigma^2, \tag{2.6}$$

where $\boldsymbol{\Gamma}_i$ is the variance matrix of β_i which is block diagonal consisting of $(1 + n_i)$ 2×2 matrices, the first being $\boldsymbol{\Sigma}_s\sigma^2$ and the other n_i being $\boldsymbol{\Sigma}_d\sigma^2$. Thus in the example above $\boldsymbol{\Gamma}_1\sigma^2$ is an 8×8 block diagonal matrix given by

$$\boldsymbol{\Gamma}_1\sigma^2 = \begin{bmatrix} \theta_1 & \theta_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_2 & \theta_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \theta_4 & \theta_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & \theta_5 & \theta_6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \theta_4 & \theta_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & \theta_5 & \theta_6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_4 & \theta_5 \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_5 & \theta_6 \end{bmatrix}$$

The records of the animals with no offspring can be easily embedded into the framework of (2.5) and (2.6). We let $\mathbf{w}_0$ be the vector of these records, $\mathbf{Z}_0$ be the identity matrix, and the variance matrix, $\boldsymbol{\Gamma}_0$, be $\theta_1\mathbf{I}$ (or $\theta_4\mathbf{I}$ by (2.4)). If we now combine all the $\mathbf{w}_i$ ($i = 0, \cdots, s$) into a vector $\mathbf{w}$ and similarly combine $\mathbf{X}_i$ and β_i into $\mathbf{X}$ and β respectively, (2.5) and (2.6) can be written as

$$\mathbf{w} = \mathbf{X}\alpha + \mathbf{Z}\beta + \mathbf{e} \text{ and } \mathbf{V} = (\mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}' + \mathbf{I})\sigma^2, \tag{2.7}$$

where $\mathbf{Z}$ is the direct sum of $\mathbf{Z}_i$ (Searle 1966, p. 213) and $\boldsymbol{\Gamma}$ is a block diagonal matrix with elements $\boldsymbol{\Gamma}_i$. In the next section we make use of this formulation for the variance matrix, showing how the fixed effects and genetic parameters might be estimated.

3. Estimation of Parameters

In this section we discuss the ML estimation of the parameters σ^2_P, σ^2_A, σ^2_K and α. We will give ML estimating equations [(3.7)–(3.9)] and indicate how they might be solved iteratively using second differentials [(3.14)–(3.16)]. We also give more practical alternative forms [(3.10)–(3.13), (3.17) and (3.18)] of these quantities using terms which can be accumulated over each sire family. The formulae involved are rather complex so we relegate definition of the terms used in the alternative forms and discussion of the relationship between the two forms to the appendix.

The method follows closely one discussed by Patterson and Thompson (1971), the essential difference being that in this paper we find it useful, since we have two generations, to work with 2×2 matrices, whereas Patterson and Thompson (1971) were able to work with scalars. They worked with a variance matrix of the form $\mathbf{V} = \mathbf{H}\sigma^2$ where $\mathbf{H}$ was a linear function of unknown parameters. In order to use their results we form

$$\gamma_A = \sigma^2_A/\sigma^2 \quad \text{and} \quad \gamma_K = \sigma^2_K/\sigma^2$$

and express the estimation procedure in terms of γ_A, γ_K and σ^2. Once these parameters have been estimated it is easy to find estimates of the genetic parameters σ^2_P, σ^2_A and σ^2_K. Obviously, the two parameterizations give equivalent results. We use the γ parameterization as it is easier to implement computationally, although results in terms of the natural genetic parameterization are perhaps easier to interpret.

We now introduce notation so that the matrix, $\mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}' + \mathbf{I}$, is linear in the unknown parameters and the results of Patterson and Thompson (1971) can be easily used. The matrix

Γ is block diagonal, made up of blocks of Σ_s and Σ_d, and these can be written in terms of γ_A and γ_K as

$$\Sigma_s = \mathbf{D}_{sA}\gamma_A + \mathbf{D}_{sK}\gamma_K, \tag{3.1}$$

$$\Sigma_d = \mathbf{D}_{dA}\gamma_A + \mathbf{D}_{dK}\gamma_K, \tag{3.2}$$

where

$$\mathbf{D}_{sA} = \mathbf{D}_{dA} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{4} \end{pmatrix}, \qquad \mathbf{D}_{sK} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \mathbf{D}_{dK} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We can then write the variance matrix of $\mathbf{w}$ in the form $\mathbf{H}\sigma^2 = (\mathbf{Z}\mathbf{D}_A\mathbf{Z}'\gamma_A + \mathbf{Z}\mathbf{D}_K\mathbf{Z}'\gamma_K + \mathbf{I})\sigma^2$, where $\mathbf{D}_j$ is formed from $\mathbf{D}_{sj}$ and $\mathbf{D}_{dj}$ in the same way as Γ is formed from Σ_s and Σ_d ($j = A, K$). The differential of Γ with respect to γ_j is now $\mathbf{D}_j$ and this fact is used in deriving some of the results below.

Following Patterson and Thompson (1971) we maximize the log likelihood of error contrasts, L', to estimate the variance parameters, σ^2, γ_A and γ_K, arguing that other contrasts, i.e. contrasts between fixed effects, give no information on the variance parameters. In order to present the estimating equations concisely, we let

$$\mathbf{S} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}', \tag{3.3}$$

$$\tilde{\beta} = \mathbf{W}^{-1}\mathbf{Z}'\mathbf{S}\mathbf{w}, \tag{3.4}$$

$$\mathbf{U} = \Gamma^{-1} - \Gamma^{-1}\mathbf{W}^{-1}\Gamma^{-1}, \tag{3.5}$$

$$\mathbf{W} = \mathbf{Z}'\mathbf{S}\mathbf{Z} + \Gamma^{-1}, \tag{3.6}$$

and n be the total number of observations and suppose $\mathbf{X}$ has rank t. The equations of estimation for the variance parameters are then (by extending slightly the results of Patterson and Thompson [equations (15)–(32)] or from similar results in Harville (1977))

$$\frac{\partial L'}{\partial \gamma_j} = -\tfrac{1}{2}\operatorname{tr}(\mathbf{U}\mathbf{D}_j) + (\tilde{\beta}'\Gamma^{-1}\mathbf{D}_j\Gamma^{-1}\tilde{\beta})/2\sigma^2 = 0 \; (j = A, K), \tag{3.7}$$

$$\frac{\partial L'}{\partial \sigma^2} = -\tfrac{1}{2}(n - t)/\sigma^2 + (\mathbf{w}'\mathbf{S}\mathbf{w} - \mathbf{w}'\mathbf{S}\mathbf{Z}\tilde{\beta})/2\sigma^4 = 0. \tag{3.8}$$

Once the variance parameters have been estimated we use weighted least squares to estimate α i.e. $\hat{\alpha}$ satisfies

$$(\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})\alpha = \mathbf{X}'\mathbf{H}^{-1}\mathbf{w}. \tag{3.9}$$

This formulation can be useful in the context Patterson and Thompson discussed, namely incomplete block designs, but not usually in our context since it requires the inversion of the matrix $\mathbf{W}$ which is of size $s' + s + D' + D$. Instead we show in the appendix that we can express the results in an alternative form that allows us to accumulate the relevant terms sequentially and only invert several matrices of size 2, and one of size t, the number of fixed effects. In particular we show that

$$\hat{\alpha} = \mathbf{Q}(\mathbf{B} - \mathbf{C})_{12}, \tag{3.10}$$

$$\operatorname{tr}(\mathbf{U}\mathbf{D}_j) = e_j - \operatorname{tr}(\mathbf{Q}(\mathbf{E}_j)_{11}), \tag{3.11}$$

$$\tilde{\beta}'\Gamma^{-1}\mathbf{D}_j\Gamma^{-1}\tilde{\beta} = \alpha'_*\mathbf{E}_j\alpha_*, \tag{3.12}$$

and

$$\mathbf{w}'\mathbf{S}\mathbf{w} - \mathbf{w}'\mathbf{S}\mathbf{Z}\tilde{\beta} = (\mathbf{B} - \mathbf{C})_{22} - (\mathbf{B} - \mathbf{C})_{21}\hat{\alpha}, \tag{3.13}$$

where $\tilde{\beta}$, $\mathbf{C}$, $\mathbf{E}_j$, e_j, $\mathbf{Q}$, α_* and the suffix notation are defined in the appendix (equations (A7), (A8), (A9), (A11), (A14), (A15) and (A13) respectively). Details of the proofs of (3.11)-(3.13) are available from the author.

However, in general, we cannot solve (3.8) and (3.9) explicitly for the variance parameters and need an iterative scheme. One such scheme is Fisher's scoring technique, (Kendall and Stuart 1967, pp. 48–53) which is based on the information matrix. This matrix is also useful in finding asymptotic variances for the estimated variance parameters, but for some saving in computation we suggest using the second differentials of L' rather than their expected values which form the information matrix. These second differentials of L' are, by extension of the results of Patterson and Thompson (1971) or Harville (1977),

$$\frac{\partial^2 L'}{\partial \gamma_j \partial \gamma_l} = \tfrac{1}{2}\text{tr}\,(\mathbf{U}\mathbf{D}_j\mathbf{U}\mathbf{D}_l) - (\tilde{\beta}'\boldsymbol{\Gamma}^{-1}\mathbf{D}_j\mathbf{U}\mathbf{D}_l\boldsymbol{\Gamma}^{-1}\tilde{\beta})/\sigma^2 \qquad (j, l = A, K), \tag{3.14}$$

$$\frac{\partial^2 L'}{\partial \gamma_j\, \partial \sigma^2} = -(\tilde{\beta}'\boldsymbol{\Gamma}^{-1}\mathbf{D}_j\boldsymbol{\Gamma}^{-1}\tilde{\beta})/2\sigma^4 \qquad (j = A, K), \tag{3.15}$$

$$\frac{\partial^2 L'}{\partial \sigma^2\, \partial \sigma^2} = (n - t)/2\sigma^4 - (\mathbf{w}'\mathbf{S}\mathbf{w} - \mathbf{w}'\mathbf{S}\mathbf{Z}\tilde{\beta})/\sigma^6. \tag{3.16}$$

Again, alternative forms are computationally more feasible and the terms in (3.14)–(3.16) are better derived using (3.12) (3.13) and

$$\text{tr}\,(\mathbf{U}\mathbf{D}_j\mathbf{U}\mathbf{D}_l) = f_{jl} - \text{tr}\,((\mathbf{F}_{jl})_{11}\mathbf{Q}) + \text{tr}\,((\mathbf{E}_j)_{11}\mathbf{Q}(\mathbf{E}_l)_{11}\mathbf{Q}) \tag{3.17}$$

and

$$\tilde{\beta}'\boldsymbol{\Gamma}^{-1}\mathbf{D}_j\mathbf{U}\mathbf{D}_l\boldsymbol{\Gamma}^{-1}\tilde{\beta} = \tfrac{1}{2}\alpha'_*\mathbf{F}_{jl}\alpha_* - \alpha'_*\mathbf{E}_j\begin{pmatrix}\mathbf{Q} & 0\\ 0 & 0\end{pmatrix}\mathbf{E}_l\alpha_*, \tag{3.18}$$

where $\mathbf{F}_{jl}$ and f_{jl} are defined in the appendix (equations (A10) and (A12)).

In one important case the expected values of the second differentials are more difficult to calculate than the acutal values. This occurs when the individuals to have offspring are selected on their own records. The likelihood for this case can be written in terms of the likelihood of parental records and the likelihood of the offspring records given the parental records. Using this result it can be shown that the likelihood in this case is the same as when parents are selected at random. Kempthorne and von Krosigk (in Henderson, Kempthorne, Searle and von Krosigk 1959) have used a similar argument. Hence (3.3) and (3.4) could again be used to estimate the variance parameters. When parents are selected part of the information matrix can be worked out conditional on the parental values (Curnow 1961, and Thompson 1973, 1976), and whilst it was relatively straightforward to use the expected values in the papers referred to, this author finds it much simpler for the model in this paper to use the observed second differentials. Kempthorne and von Krosigk have, on the other hand, suggested (in Henderson *et al.* 1959) that the information matrix is easier to calculate than the second differentials, but they do not give a formula.

In the author's experience, the amount of attention paid to the information matrix in the simple hierarchical design on the offspring suggests that further approximations to the information matrix are acceptable in practice. If the actual data are not too unbalanced (both in the sense of numbers associated with the fixed effects, and the numbers of offspring per dam and sire), then the corresponding formulae in the balanced case (Thompson [1976]) might provide an acceptable approximation. One might use the means for the number of dams per sire and offspring per dam (d and n in Thompson's formulae) and adjust the degrees of freedom for the number of fixed effects estimated.

To summarize, one possible way of implementing the ML procedure outlined in this section is

(i) Obtain initial estimates for the variance parameters.
(ii) Accumulate (A7)–(A12) over sire families, using (A17) to find $(\mathbf{Z}'_i\mathbf{Z}_i + \mathbf{\Gamma}^{-1}{}_i)^{-1}$.
(iii) Find (A13)–(A15).
(iv) Evaluate first and second differentials [(3.7), (3.8), (3.14), (3.15) and (3.16)] using (3.10), (3.11), (3.12), (3.17) and (3.18).
(v) Find new estimates of the variance parameters.
(vi) Repeat from (ii) until the estimates satisfy (3.7) and (3.8).

4. Discussion

The ML method of estimation of the previous section can be modified fairly easily to deal with various other circumstances, for example if a different model is assumed or if observations are only available on one sex. We now outline some of these modifications.

If a maternal environmental covariance is assumed this requires $\theta_5 = \frac{1}{2}\sigma^2{}_A + \sigma^2{}_M$ in (2.4) and the formation of two differential matrices

$$\mathbf{D}_{sM} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{D}_{dM} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

for example (3.1) becomes

$$\mathbf{\Gamma}_s = \mathbf{D}_{sA}\gamma_A + \mathbf{D}_{sK}\gamma_K + D_{sM}\gamma_M$$

where $\gamma_M = \sigma^2{}_M/\sigma^2{}_P$ and terms analogous to (3.7), (3.15) and (3.16) can be formed for the maternal component.

In some cases observations are not available on male or female parents. This necessitates the deletion of rows and columns from the respective 2×2 matrices. For instance if male parents are not measured then

$$\mathbf{D}_{sA} = (\tfrac{1}{4}), \qquad \mathbf{D}_{sK} = (0), \qquad \mathbf{A}_{oo} = (n_{io}) \tag{4.1}$$

and

$$\mathbf{A}'_{oj} = (0 \;\; n_{ij}) \tag{4.2}$$

($\mathbf{A}_{oo}$ and $\mathbf{A}'_{oj}$ are given in the appendix). If female parents are not measured then

$$\mathbf{D}_{dA} = (\tfrac{1}{4}), \qquad \mathbf{D}_{dK} = (1), \qquad \mathbf{A}_{jj} = (n_{ij}), \qquad A_{jK} = (0) \tag{4.3}$$

and

$$\mathbf{A}_{jo} = (0 \;\; n_{ij}). \tag{4.4}$$

If neither male nor female parents are measured then (4.1) and (4.3) hold and $\mathbf{A}_{jo} = (n_{ij})$, and $\mathbf{\Gamma}$ is a diagonal matrix. In this case, when observations are available only on progeny, a reparameterization in terms of half and full sib covariances suggests itself, and there are well established 'analysis of variance' techniques of equating sums of squares to their expectation (Searle 1971) to estimate these components. It is difficult to give a useful rule of thumb to say when the more complex ML procedure is worth following since this naturally depends on the degree of inbalance in the family structure, on the orthogonality of the fixed and random effects, and on the values of the variance parameters. When the family structure is balanced

and the fixed effects are orthogonal to the random effects the two methods give the same estimates.

We have taken advantage of the special structure of the data by partitioning the model into two parts $\mathbf{X}$ and $\mathbf{Z}$ corresponding to the fixed and random effects, and we have absorbed the terms relating to $\mathbf{Z}$ in order to give a convenient form for the computations. This method is not restricted to just this genetic model but could be useful in any other variance component problem where $\mathbf{Z}'\mathbf{Z} + \mathbf{\Gamma}^{-1}$ can be split up into block diagonal form. We can easily extend the results to deal with the case when both $\mathbf{X}$ and $\mathbf{Z}$ contain fixed and random effects and details are available from the author.

It is perhaps of interest to give an example of how this approach can prove useful. Suppose data are available on several cereal varieties (say 20) grown for a number of years (say ten) at a number of sites (say five). Not all varieties are grown in all years or at all sites. Suppose we want to estimate variety effects, and year, centre, year by centre, year by variety and year by centre components. Since there are more varieties than centres or years we associate variety (fixed), variety by year (random) and variety by centre (random) effects with $\mathbf{Z}$ and centre, year and centre by year (all random) effects with $\mathbf{X}$. Terms (A7)–(A12) can be formed separately for each variety and accumulated sequentially.

Acknowledgements

I am very grateful to Oscar Kempthorne and a referee for very thorough reviews and to W. G. Hill for useful comments.

L'Estimation De L'Heritabilite A Partir De Donnees Non Equilibrees
I—Observations possibles sur les parents et leurs enfants

Résumé

On discute l'estimation du maximum de vraisemblance (ML) de l'héritabilité dans le cas où on dispose de données sur, à la fois les parents et leurs enfants, et où les données sur les enfants ont une structure emboîtée (hiérarchique). On développe des modes de calcul qui sont utiles quand les familles de germains ou de demi-frères ou soeurs présentent des dimensions différentes. Des extensions à d'autres situations d'analyse des composants de la variance sont discutées.

References

Curnow, R. N. (1961). The estimation of repeatability and heritability from records subject to culling. *Biometrics 17,* 553–566.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association 72,* to appear.

Hemmerle, W. J. and Hartley, H. O. (1973). Computing maximum likelihood estimates for the mixed A.O.V. model using the W transformation. *Technometrics 15,* 819–831.

Henderson, C. R. (1963). Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding.* W. D. Hanson and H. F. Robinson (eds.), National Academy of Sciences, National Research Council, Washington, D. C., Publication 982, 141–163.

Henderson, C. R., Kempthorne, O., Searle, S. R., and Von Krosigk, C. M. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics 15,* 192–218.

Hill, W. G. and Nicholas, F. W. (1974). Estimation of heritability by both regression of offspring on parent and intra-class correlation of sibs in one experiment. *Biometrics 30,* 447–468.

Kempthorne, O. and Tandon, O. B. (1952). The estimation of heritability by regression of offspring on parent. *Biometrics 9,* 90–100.

Kendall, M. G. and Stuart, A. (1967). *The Advanced Theory of Statistics,* Vol. II, 2nd ed., Griffin, London.
Ollivier, L. (1974). La regression parent-descendant dans le cas de descendances subdivisees en familles detaille inegale. *Biometrics 30,* 59–66.
Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika 58,* 545–554.
Robertson, A. (1962). Weighting in the estimation of variance components in the unbalanced single classification. *Biometrics 18,* 413–417.
Searle, S. R. (1966). *Matrix Algebra for the Biological Sciences. Wiley, New York.*
Searle, S. R. (1971). Linear Models. Wiley, New York.
Thompson, R. (1973). The estimation of variance and covariance components with an application when records are subject to culling. *Biometrics 29,* 527–550.
Thompson, R. (1976). Design of experiments to estimate heritability when observations are available on parents and offspring. *Biometrics 32,* 283–304.

Received March 1976; *Revised October* 1976

Appendix

In this appendix we first discuss alternate forms for quantities arising in the estimation process, [(3.7)–(3.9) and (3.14)–(3.16)]. We then show how the inverse of $\mathbf{Z}'_i\mathbf{Z}_i + \Gamma^{-1}{}_i$ can be formed in terms of 2×2 matrices.

Discussion of results of Section 3

The key result to note is that $\hat{\alpha}$ and $\tilde{\beta}$ as given by (3.7) and (3.4) can be found from

$$\mathbf{X}'\mathbf{X}\alpha + \mathbf{X}'\mathbf{Z}\,\tilde{\beta} = \mathbf{X}'\mathbf{w} \tag{A1}$$

$$\mathbf{Z}'\mathbf{X}\alpha + (\mathbf{Z}'\mathbf{Z} + \Gamma^{-1})\tilde{\beta} = \mathbf{Z}'\mathbf{w} \tag{A2}$$

as given by Henderson in Henderson, *et al.* (1959). Henderson (1973) has emphasized that the $\tilde{\beta}$ values have a genetic interpretation in terms of the breeding values of the individuals.

As $\mathbf{Z}'\mathbf{Z} + \Gamma^{-1}$ is a block diagonal matrix, we can estimate α by eliminating the set of $\tilde{\beta}$'s for each sire group in turn using (A2) to give

$$\sum_i (\mathbf{X}_i'\mathbf{X}_i - \mathbf{X}_i'\mathbf{Z}_i(\mathbf{Z}_i'\mathbf{Z}_i + \Gamma_i{}^{-1})^{-1}\mathbf{Z}_i\mathbf{X}_i)\hat{\alpha} = \sum_i (\mathbf{X}_i'(\mathbf{I} - \mathbf{Z}_i(\mathbf{Z}_i'\mathbf{Z}_i + \Gamma_i{}^{-1})^{-1}\mathbf{Z}_i'\mathbf{w}_i) \tag{A3}$$

which is equivalent to (3.9).

We also require an alternative form for $\mathbf{U}$

$$\mathbf{U} = \mathbf{U}^* - \Gamma^{-1}(\mathbf{Z}'\mathbf{Z} + \Gamma^{-1})^{-1}\mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \Gamma^{-1})^{-1}\Gamma^{-1} \tag{A4}$$

where

$$\mathbf{U}^* = \Gamma^{*-1} - \Gamma^{-1}(\mathbf{Z}'\mathbf{Z} + \Gamma^{-1})^{-1}\Gamma^{-1}, \tag{A5}$$

using the expansion for $\mathbf{W}^{-1}$ given by equation (42) in Patterson and Thompson (1971).

Results (A3)–(A5) depend on the inverse of $(\mathbf{Z}'_i\mathbf{Z}_i + \Gamma^{-1}{}_i)$ which is of size $2 \times (1 + n_i)$. This matrix is highly structured, and we show below how to express its inverse in terms of 2×2 matrices.

We now define several matrices which can be accumulated over sire families so that we may rewrite the estimating equations. We first form a $n \times (t + 1)$ matrix, $\mathbf{Y}$, from $\mathbf{X}$ and $\mathbf{w}$ so that

$$\mathbf{Y} = (\mathbf{X}\ \mathbf{w}). \tag{A6}$$

This enables the matrices to be defined more compactly and is suggested by the work of Hemmerle and Hartley (1973). We now let $\mathbf{B}$, $\mathbf{C}$, $\mathbf{E}_j$, $\mathbf{F}_{jl}$ $(j, l = A, K)$ be $(t + 1) \times (t + 1)$ matrices given by

$$\mathbf{B} = \mathbf{Y}'\mathbf{Y}, \tag{A7}$$

$$\mathbf{C} = \mathbf{Y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \Gamma^{-1})^{-1}\mathbf{Z}'\mathbf{Y}, \tag{A8}$$

$$\mathbf{E}_j = \mathbf{Y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \mathbf{\Gamma}^{-1})^{-1}\mathbf{\Gamma}^{-1}\mathbf{D}_j\mathbf{\Gamma}^{-1}(\mathbf{Z}'\mathbf{Z} + \mathbf{\Gamma}^{-1})^{-1}\mathbf{Z}'\mathbf{Y}, \tag{A9}$$

$$\mathbf{F}_{jl} = \mathbf{G}_{jl} + \mathbf{G}'_{jl}, \tag{A10}$$

where

$$\mathbf{G}_{jl} = \mathbf{Y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \mathbf{\Gamma}^{-1})^{-1}\mathbf{\Gamma}^{-1}\mathbf{D}_j(\mathbf{\Gamma}^{-1} - \mathbf{\Gamma}^{-1}(\mathbf{Z}'\mathbf{Z} + \mathbf{\Gamma}^{-1})^{-1}\mathbf{\Gamma}^{-1})\mathbf{D}_l\mathbf{\Gamma}^{-1}(\mathbf{Z}'\mathbf{Z} + \mathbf{\Gamma}^{-1})^{-1}\mathbf{Z}'\mathbf{Y}.$$

The matrices $\mathbf{B}$, $\mathbf{C}$, $\mathbf{E}_j$, $\mathbf{F}_{jl}$ can be accumulated sequentially over sire families, i.e. $\mathbf{B} = \Sigma_i \mathbf{Y}'_i \mathbf{Y}_i$, and since they are all symmetric we need only compute the upper (or lower) triangular part of each one. $\mathbf{E}_j$ and $\mathbf{F}_{jl}$ can be thought of as weighted sums of squares and products of the variates $\mathbf{\Gamma}^{-1}(\mathbf{Z}'\mathbf{Z} + \mathbf{\Gamma}^{-1})^{-1}\mathbf{Z}'\mathbf{Y}$.

We also need

$$e_j = \text{tr}(\mathbf{D}_j\mathbf{U}^*) \tag{A11}$$

and

$$f_{jl} = \text{tr}(\mathbf{D}_j\mathbf{U}^*\mathbf{D}_l\mathbf{U}^*) \tag{A12}$$

where $\mathbf{U}^*$ is given by (A5). Again e_j and f_{jl} can be accumulated over sire families.

Now we partition the matrices defined by (A6) to (A10) according to whether they are associated with $\mathbf{X}$ or $\mathbf{w}$, i.e.

$$(\mathbf{B})_{11} = \mathbf{X}'\mathbf{X}, \ (\mathbf{B})_{12} = \mathbf{X}'\mathbf{w}, \tag{A13}$$

$$(\mathbf{B})_{21} = \mathbf{w}'\mathbf{X}, \ (\mathbf{B})_{22} = \mathbf{w}'\mathbf{w}.$$

Now $(\mathbf{B} - \mathbf{C})_{11}$ is the matrix on the left hand side of (A3) and hence is $\mathbf{X}'\mathbf{H}^{-1}\mathbf{X}$ and correspondingly $(\mathbf{B} - \mathbf{C})_{12}$ is the right hand side of (A3). If we let

$$\mathbf{Q} = ((\mathbf{B} - \mathbf{C})_{11})^{-1} \tag{A14}$$

then

$$\hat{\alpha} = \mathbf{Q}(\mathbf{B} - \mathbf{C})_{12}.$$

Using (A6), $(\mathbf{w} - \mathbf{X}\hat{\alpha}) = \mathbf{Y}\alpha_*$ where

$$\alpha_* = (-\hat{\alpha} \quad 1)'. \tag{A15}$$

We now show how the inverse of $\mathbf{Z}'_i\mathbf{Z}_i + \mathbf{\Gamma}^{-1}{}_i$ can be expressed in terms of 2×2 matrices. We use the matrix identity for partitioned matrices

$$\begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{M}_{11.} & -\mathbf{M}_{11.}\mathbf{M}_{12}\mathbf{M}_{22}{}^{-1} \\ -\mathbf{M}_{22}{}^{-1}\mathbf{M}_{21}\mathbf{M}_{11.} & \mathbf{M}_{22}{}^{-1} + \mathbf{M}_{22}{}^{-1}\mathbf{M}_{21}\mathbf{M}_{11.}\mathbf{M}_{12}\mathbf{M}_{22}{}^{-1} \end{pmatrix} \tag{A16}$$

where

$$\mathbf{M}_{11} = (\mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21})^{-1}.$$

The matrix $(\mathbf{Z}'_i\mathbf{Z}_i + \mathbf{\Gamma}^{-1}{}_i)$ can be partitioned into a matrix of the form

$$\begin{pmatrix} \mathbf{A}_{oo} & \mathbf{A}_{o1} & \cdots & \mathbf{A}_{on_i} \\ \vdots & & & \vdots \\ \mathbf{A}_{n_io} & & \cdots & \mathbf{A}_{n_in_i} \end{pmatrix}$$

where $\mathbf{A}$'s are 2×2 matrices

and

$$\mathbf{A}_{oo} = \begin{pmatrix} 1 & 0 \\ 0 & n_{io} \end{pmatrix} + \mathbf{\Sigma}^{-1}{}_s, \ \mathbf{A}_{jj} = \begin{pmatrix} 1 & 0 \\ 0 & n_{ij} \end{pmatrix} + \mathbf{\Sigma}^{-1}{}_d \quad 1 \le j \le n_i,$$

$$\mathbf{A}_{j0} = \begin{pmatrix} 0 & 0 \\ 0 & n_{ij} \end{pmatrix} = \mathbf{A}'_{oj} \ 1 \le j \le n_i \quad \text{and} \quad \mathbf{A}_{jk} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} 1 \le j, k \le n_i, j \ne k,$$

where $\sum_j n_{ij} = n_{io}$ and $\mathbf{\Sigma}_s$ and $\mathbf{\Sigma}_d$ are given by (2.3).

The inverse of $\mathbf{Z}'_i\mathbf{Z}_i + \mathbf{\Gamma}^{-1}_i$ can be written in the same form with 2×2 matrices $\mathbf{K}_{jk}$ $(0 \le j, k \le n_i)$. If $(\mathbf{A}_{jk})_{22}$ is the element in the second row and column of $\mathbf{A}_{jk}$ then by partitioning $\mathbf{Z}'_i\mathbf{Z}_i + \mathbf{\Gamma}^{-1}_i$ into terms associated with sires and dams and using (A16) we find

$$\mathbf{K}^{-1}{}_{oo} = \mathbf{A}_{oo} - \begin{pmatrix} 0 & 0 \\ 0 & \sum_j n^2{}_{ij}(\mathbf{A}^{-1}{}_{jj})_{22} \end{pmatrix}$$

$$\mathbf{K}_{oj} = -\mathbf{K}_{oo}\mathbf{A}_{jo}\mathbf{A}^{-1}{}_{jj} = \mathbf{K}'_{jo} \tag{A17}$$

and

$$\mathbf{K}_{jk} = \mathbf{A}^{-1}{}_{jj}\delta_{jk} + \mathbf{A}^{-1}{}_{jj}\begin{pmatrix} 0 & 0 \\ 0 & n_{ij}n_{ik}(\mathbf{K}_{oo})_{22} \end{pmatrix}\mathbf{A}^{-1}{}_{kk}$$

where δ_{jk} is the Kronecker delta.

Reprinted from
BIOMETRICS COPYRIGHT © 1977
THE BIOMETRIC SOCIETY, Vol. 33, No. 3, September 1977

19

The estimation of heritability with unbalanced data. II. Data available on more than two generations

by

R. Thompson

BIOMETRICS *33*, 497–504
September, 1977

The Estimation of Heritability with Unbalanced Data II. Data Available on more than Two Generations

ROBIN THOMPSON

A.R.C. Unit of Statistics, University of Edinburgh, Edinburgh EH9 3JZ, Scotland

Summary

The maximum likelihood (ML) estimation of heritability is considered when data are available from one sex and for more than two generations and the progeny are non-inbred half-sibs. The covariances between relatives in different generations are derived. ML estimating equations are given using results derived in a companion paper (Thompson 1977). Some of the formulae derived depend on the inversion of matrices and a convenient tabular form for deriving these inverse matrices is given that takes advantage of their special structure. Estimation when data are available on both sexes is briefly discussed.

1. Introduction

In a companion paper (Thompson 1977) the maximum likelihood (ML) estimation of heritability when data are available on parents and progeny is discussed. In this paper the ML estimation of heritability when observations are available for more than two generations is discussed, where the progeny are assumed to be non-inbred half-sibs and observations are only available on one sex. Though theory is easier to implement than when both sexes are measured, it includes some cases of practical importance. For instance, the progeny testing of dairy bulls for milk yield, and the performance and progeny testing of bulls for beef characteristics, fit into this framework.

It is perhaps useful to outline a practical example which initiated the discussion in this paper. In Malawi the performance testing of Zebu bulls has been carried out for a number of years. The results for eight years were available for analysis. The data can be roughly split into two parts. In the first three years the bulls that were measured (for liveweight gain and food intake over a specified period) were progeny of bulls that had not been measured, but in the last five years the bulls that were measured were mainly the progeny of bulls measured in previous years. In fact for some of the bulls both the sire and the grandsire had been previously measured. Interest was expressed in the estimation of various environmental effects such as effects of different stations, of different years and of the age of the dam, and in the estimation of the additive and phenotypic variance.

Obviously, there is information on heritability both in the sib-covariances and in the parent-offspring covariances. We use an ML approach in an attempt to combine all the available information efficiently. It seems useful to discuss this ML approach since it is a simple extension of results of a previous paper (Thompson 1977).

In Section 2 we generate the covariances between relatives in different generations in

Key words: Heritability estimation; Maximum likelihood estimation; Unbalanced data; Relationship matrix.

terms of the additive genetic variance. We present the results on ML estimation in Section 3 using results given in Part I (Thompson 1977). We briefly discuss the case when both sexes are measured in Section 4.

2. *Generating Covariances Between Relatives*

In order to use the ML approach we need to know the covariances between animals. In this section we give a convenient way of generating the covariances between relatives in different generations in terms of the additive genetic variance. We assume, for simplicity, that generations are discrete so that an individual in generation i is the offspring of parents in generation $(i - 1)$, and there are N_i animals in the i *th* generation. Suppose we start in generation 0 and we order the individuals by sex, males first and then females, assuming that they are unrelated and the offspring of non-inbred parents. So $\mathbf{R}_{00}$, a matrix representing the coefficients of parentage (Kempthorne's 1957 translation of Malécot's 1948 "coefficient de parenté") is $\frac{1}{2}\mathbf{I}$. That is, the (r, s) element of $\mathbf{R}_{00}$ represents the coefficient of parentage between the r *th* and s *th* individuals. In a similar manner we can define $\mathbf{R}_{ij}$ as a $(N_i \times N_j)$ matrix representing the coefficients of parentage between individuals in the i *th* and j *th* generations, the (r, s) element of $\mathbf{R}_{ij}$ being the coefficient of parentage between the r *th* individual in generation i with the s *th* individual in generation j.

The coefficients of parentage between individuals can be written in algebraic form in terms of the coefficients of parentage of relatives. For instance, if individual A is in a later generation than individual B then the coefficient of parentage, r_{AB}, can be written as (Kempthorne 1957)

$$r_{AB} = \tfrac{1}{2}(r_{WB} + r_{XB}), \tag{2.1}$$

and if A and B are in the same generation $(A \neq B)$ then

$$r_{AB} = \tfrac{1}{4}(r_{WY} + r_{WZ} + r_{XY} + r_{XZ}) \tag{2.2}$$

where X and W are the parents of A, and Y and Z are the parents of B. Also

$$r_{AA} = \tfrac{1}{2}(1 + r_{XW}) \tag{2.3}$$

$$= (r_{WW} + r_{XW} + r_{WX} + r_{XX}) + \tfrac{1}{2}(1 - \tfrac{1}{2}r_{WW} - \tfrac{1}{2}r_{XX}). \tag{2.4}$$

In the form (2.4) the formula for r_{AA} is similar to (2.2) but with an extra term $\frac{1}{2}(1 - \frac{1}{2}r_{WW} - \frac{1}{2}r_{XX})$.

So if we have the relationships between individuals in the i *th* and k *th* generations, we can easily derive the relationships between offspring of the i *th* generation, i.e. the $i + 1$ generation, and individuals of the k *th* generation using (2.1) (if $i \geq k$). Then the matrix $\mathbf{R}_{(i+1)k}$ can be represented by

$$\mathbf{R}_{(i+1)k} = \mathbf{Z}_i \mathbf{R}_{ik} \tag{2.5}$$

where $\mathbf{Z}_i$ is a $N_{i+1} \times N_i$ matrix relating the individuals of generation $(i + 1)$ to those of generation i.

The $\mathbf{Z}_i$ matrices can be written as

$$\begin{pmatrix} \mathbf{Z}_{imm} & \mathbf{Z}_{imf} \\ \mathbf{Z}_{ifm} & \mathbf{Z}_{iff} \end{pmatrix}, \tag{2.6}$$

where the elements of $\mathbf{Z}_i$ are either 0 or $\frac{1}{2}$. The (j, k) element of $\mathbf{Z}_{imm}$ is $\frac{1}{2}$ only if the k *th* male of generation i is the father of the j *th* male of generation $(i + 1)$, and similarly the (j, k) element

of $\mathbf{Z}_{imf}$ is $\frac{1}{2}$ only if the *k th* female of generation *i* is the mother of the *j th* male of generation $(i + 1)$. Likewise the elements of $\mathbf{Z}_{ifm}$ and $\mathbf{Z}_{iff}$ can be defined in terms of the parentage of the female offspring. Hill (1974) used similar matrices in a study of the prediction of selection response with overlapping generations. He suggested that the blocks of $\mathbf{Z}$ could be thought of as the alternative pathways of genes

$$\left(\begin{array}{c|c} \text{males from males} & \text{males from females} \\ \hline \text{females from males} & \text{females from females} \end{array}\right)$$

The relationship matrix between individuals in the $(i + 1)$ generation, $\mathbf{R}_{(i+1)(i+1)}$, can be written in terms of the relationship matrix of their parents, $\mathbf{R}_{ii}$, using (2.4). Using the $\mathbf{Z}$ matrices we find

$$\mathbf{R}_{(i+1)(i+1)} = \mathbf{Z}_i\mathbf{R}_{ii}\mathbf{Z}'_i + \mathbf{D}_{ii} \tag{2.7}$$

where $\mathbf{D}_{ii}$ is a diagonal matrix containing the terms analogous to $\frac{1}{2}(1 - \frac{1}{2}r_{ww} - \frac{1}{2}r_{xx})$ in (2.4). $\mathbf{D}_{ii}$ is a function of the coefficients of parentage of the parents of the individuals and could be expressed in terms of matrix operators. However, for simplicity we assume that $r_{ww} = r_{xx} = \frac{1}{2}$, implying that the parents are not inbred and so $\mathbf{D}_{ii} = \frac{1}{4}I_{(i+1)}$, where $\mathbf{I}_i$ is an identity matrix of size $(N_i \times N_i)$. (We will drop the suffix *i* when the size of $\mathbf{I}_i$ is self-evident) and

$$\mathbf{R}_{(i+1)(i+1)} = \mathbf{Z}_i\mathbf{R}_{ii}\mathbf{Z}'_i + \tfrac{1}{4}\mathbf{I}_{(i+1)}. \tag{2.8}$$

Equations (2.5) and (2.8) are convenient matrix representations for $\mathbf{R}_{ij}$ and could be used successively to define all $\mathbf{R}_{ij}$, but they are not necessarily the best formulation for the actual computation of $\mathbf{R}_{ij}$. The matrices $\mathbf{Z}_i$ are very sparse, each row of the four blocks of $\mathbf{Z}_i$ having only one non-zero element, and the efficient computation of $\mathbf{R}_{ij}$ should take account of this.

If we assume that the individuals in the zero *th* generation are unrelated and not inbred, then the relationship matrix $\mathbf{R}$ for the first three generations is, using (2.5) and (2.8)

$$\begin{aligned} \mathbf{R} &= \begin{bmatrix} \mathbf{R}_{00} & \mathbf{R}_{01} & \mathbf{R}_{02} \\ \mathbf{R}_{10} & \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{20} & \mathbf{R}_{21} & \mathbf{R}_{22} \end{bmatrix} \\ &= \tfrac{1}{2}\begin{bmatrix} \mathbf{I}_0 & \mathbf{I}_0\mathbf{Z}'_0 & \mathbf{I}_0\mathbf{Z}'_0\mathbf{Z}'_1 \\ \mathbf{Z}_0\mathbf{I}_0 & \mathbf{Z}_0\mathbf{I}_0\mathbf{Z}'_0 + \frac{1}{2}\mathbf{I}_1 & \mathbf{Z}_0\mathbf{I}_0\mathbf{Z}'_0\mathbf{Z}'_1 + \frac{1}{2}\mathbf{I}_1\mathbf{Z}'_1 \\ \mathbf{Z}_1\mathbf{Z}_0\mathbf{I}_0 & \mathbf{Z}_1\mathbf{Z}_0\mathbf{I}_0\mathbf{Z}'_0 + \frac{1}{2}\mathbf{Z}_1\mathbf{I}_1 & \mathbf{Z}_1\mathbf{Z}_0\mathbf{I}_0\mathbf{Z}'_0\mathbf{Z}'_1 + \frac{1}{2}\mathbf{I}_2 \end{bmatrix} \end{aligned} \tag{2.9}$$

This can be written as (using (2.5) and (2.7))

$$\boldsymbol{R} = \tfrac{1}{2}\begin{bmatrix} \mathbf{I}_0 & 0 & 0 \\ \mathbf{Z}_0 & \mathbf{I}_1 & 0 \\ \mathbf{Z}_1\mathbf{Z}_0 & \mathbf{Z}_1 & \mathbf{I}_2 \end{bmatrix}\begin{bmatrix} \mathbf{I}_0 & 0 & 0 \\ 0 & \frac{1}{2}\mathbf{I}_1 & 0 \\ 0 & 0 & \frac{1}{2}\mathbf{I}_2 \end{bmatrix}\begin{bmatrix} \mathbf{I}_0 & \mathbf{Z}'_0 & \mathbf{Z}'_0\mathbf{Z}'_1 \\ 0 & \mathbf{I}_1 & \mathbf{Z}'_1 \\ 0 & 0 & \mathbf{I}_2 \end{bmatrix}, \tag{2.10}$$

where 0 is a matrix with all elements zero, or in general as $\frac{1}{2}\mathbf{ZDZ}'$ where $\mathbf{Z}$ is a lower triangular matrix with the (i, j) sub-matrix of $\mathbf{Z}'$ given by

$$\begin{aligned} \mathbf{Z}'_{ij} &= \prod_{k=i}^{j-1} \mathbf{Z}'_k \qquad j > i, \\ \mathbf{Z}'_{ij} &= \mathbf{I}_i \qquad\qquad j = i, \\ \mathbf{Z}'_{ij} &= 0 \qquad\qquad j < i, \end{aligned} \tag{2.11}$$

$(i, j = 0, \cdots, t)$ and $\mathbf{D}$ is a block diagonal matrix with i *th* diagonal matrix

$$\begin{aligned} \mathbf{D}_i &= \mathbf{I}_i \qquad i = 0, \\ \mathbf{D}_i &= \tfrac{1}{2}\mathbf{I}_i \qquad i > 0. \end{aligned} \tag{2.12}$$

The variance matrix of the observations $(\mathbf{y}_0, \mathbf{y}_1, \cdots, \mathbf{y}_t)$ can then be written as

$$\mathbf{V} = \mathbf{ZDZ}' \sigma^2{}_A + \mathbf{I}\sigma^2{}_e \tag{2.13}$$

or

$$\mathbf{V} = (\mathbf{Z}\Gamma\mathbf{Z}' + \mathbf{I})\sigma^2{}_e$$

where

$$\Gamma = \mathbf{D}(\sigma^2{}_A/\sigma^2{}_e),$$

if we include only additive genetic variance $(\sigma^2{}_A)$ and environmental variance $(\sigma^2{}_e)$ in the formula for the variance. The $\mathbf{ZDZ}'\sigma^2{}_A$ term arises since the genetic covariance between relatives is twice the coefficient of parentage times the additive variance (Kempthorne 1957). For instance, the variance matrix for the second generation, $\mathbf{V}_{22}$, is

$$(\mathbf{Z}_1\mathbf{Z}_0\mathbf{Z}'{}_0\mathbf{Z}'{}_1 + \tfrac{1}{2}\mathbf{Z}_1\mathbf{Z}'{}_1 + \tfrac{1}{2}\mathbf{I})\sigma^2{}_A + \mathbf{I}\sigma^2{}_e. \tag{2.14}$$

$\mathbf{Z}_1\mathbf{Z}_0$ is a matrix relating the pathways of genes from individuals in the 0 *th* generation to individuals in the second generation. If the individuals are not inbred there are 4 elements of $\frac{1}{4}$ in each row of $\mathbf{Z}_1\mathbf{Z}_0$ and the other elements are 0. The variance of an individual in the second generation is then

$$\begin{aligned} (4 \times (\tfrac{1}{4})^2 + 2 \times \tfrac{1}{2}(\tfrac{1}{2})^2 + \tfrac{1}{2})\sigma^2{}_A + \sigma^2{}_e &= (\tfrac{1}{4} + \tfrac{1}{4} + \tfrac{1}{2})\sigma^2{}_A + \sigma^2{}_e \\ &= \sigma^2{}_A + \sigma^2{}_e, \text{ as one might expect.} \end{aligned} \tag{2.15}$$

The $\frac{1}{2}\mathbf{I}\sigma^2{}_A$ term in (2.14) can be interpreted as arising from the sampling of the parental gametes to form offspring. Similarly the $\frac{1}{2}\mathbf{Z}_1\mathbf{Z}'_1\sigma^2{}_A$ term can be thought of as an averaging out of the effect of sampling the grand-parental gametes to form parents. More generally, the contribution from the sampling of the i *th* generation to the variance of an individual in the j *th* generation is $(\frac{1}{2})^{j-i}\sigma^2{}_A (i < j)$. The total contribution from the sampling in all previous generations is then $\sum(\frac{1}{2})^{j-i}\sigma^2{}_A = \sigma^2{}_A$, again as one might expect.

Henderson (1976) has recently derived a formula similar to (2.10) for the case of overlapping generations. In his formula he uses $\mathbf{A}$, defined as the numerator relationship matrix, which is twice the coefficient of the parentage matrix, $\mathbf{R}$. He first gives a recursive method for computing $\mathbf{A}$ and then finds a lower triangular matrix, $\mathbf{L}$, such that $\mathbf{LL}' = \mathbf{A}$. He then expresses $\mathbf{L}$ as $\mathbf{TD}$ where $\mathbf{T}$ is a lower triangular matrix and $\mathbf{D}$ is a diagonal matrix. Hence $\mathbf{A} = \mathbf{TD}^2\mathbf{T}'$ and Henderson's $\mathbf{T}$ and $\mathbf{D}^2$ are equivalent to $\mathbf{Z}$ and $\mathbf{D}$ in this paper. Henderson shows further that $\mathbf{T}^{-1}$ and $\mathbf{A}^{-1}$ have a simple form, as does $\mathbf{Z}^{-1}$. For example if $t = 2$,

$$\mathbf{Z}^{-1} = \begin{pmatrix} \mathbf{I}_0 & 0 & 0 \\ -\mathbf{Z}_0 & \mathbf{I}_1 & 0 \\ 0 & -\mathbf{Z}_1 & \mathbf{I}_2 \end{pmatrix}$$

or, to generalize, all the blocks, $\mathbf{Z}^{-1}{}_{ij}$, of $\mathbf{Z}^{-1}$ are null matrices except

$$\begin{aligned} \mathbf{Z}^{-1}{}_{ij} &= \mathbf{I}_i \qquad (i = 0, \cdots, t), \\ \mathbf{Z}^{-1}{}_{(i+1)i} &= -\mathbf{Z}_i \qquad (i = 0, \cdots, t-1). \end{aligned}$$

The author believes that the derivation presented here is simpler and more informative than Henderson's. Firstly the terms of the diagonal matrix $\mathbf{D}$ (Henderson's $\mathbf{D}^2$) arise naturally from writing the coefficient of parentage of an individual as (2.4) or (3.1). Secondly, $\mathbf{Z}$ is easy to find since the (i, j) *th* element of $\mathbf{Z}$ is a weighted sum of the contributions of individual i to the pedigree of the individual j. The weight is a half if i is a parent of j, a quarter if i is a grandparent of j, an eighth if i is a great-grandparent of j, etc. Finally the formulation of this section has been useful to the author in understanding some of the intuitive arguments used by Hill (1971) when considering the estimation of realized heritability.

3. Measurements on One Sex

The results of the previous section will be used now in a discussion of the case when observations are only available on one sex and the progeny are assumed to be non-inbred half-sibs.

If we take the sires on whom no measurements are available as being in generation 0 then the measurements on their offspring will be, in the notation of the previous section, $\mathbf{y}_1$, and the measurements on their grandsons and great-grandsons, $\mathbf{y}_2$ and $\mathbf{y}_3$. To define the variance matrix we need to know therefore $\mathbf{Z}_0$, $\mathbf{Z}_1$, $\mathbf{Z}_2$ and $\mathbf{R}_{00}$, the coefficient of parentage matrix in the 0 *th* generation. We assume $\mathbf{R}_{00} = \frac{1}{2}\mathbf{I}$ and, for the purpose of illustration, that the dams of all the bulls were unrelated to each other and to their mates. In fact, for the Malawi example discussed in the introduction some dams had more than one measured son.

The complete coefficient of parentage matrix, $\mathbf{R}$, can be derived using the arguments of Section 2 either by putting $\mathbf{Z}_{lmf} = \frac{1}{2}\mathbf{I}$ and collecting together terms from the females, or by writing

$$r_{AA} = \tfrac{1}{4}r_{WW} + \tfrac{1}{2}(1 - \tfrac{1}{2}r_{WW}) \tag{3.1}$$

from (2.4) using the fact that the sires are not related to the dams. We find that $\mathbf{R}$ is again of the form $\frac{1}{2}\mathbf{ZDZ}'$ where now

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_{0mm} & \mathbf{I}_1 & 0 & 0 \\ \mathbf{Z}_{1mm}\mathbf{Z}_{0mm} & \mathbf{Z}_{1mm} & \mathbf{I}_2 & 0 \\ \mathbf{Z}_{2mm}\mathbf{Z}_{1mm}\mathbf{Z}_{0mm} & \mathbf{Z}_{2mm}\mathbf{Z}_{1mm} & \mathbf{Z}_{2mm} & \mathbf{I}_3 \end{pmatrix} \tag{3.2}$$

$$= (\mathbf{Z}^*_0 \quad \mathbf{Z}^*_1 \quad \mathbf{Z}^*_2 \quad \mathbf{Z}^*_3),$$

and

$$\mathbf{D} = \begin{pmatrix} \mathbf{I}_0 & 0 & 0 & 0 \\ 0 & \frac{3}{4}\mathbf{I}_1 & 0 & 0 \\ 0 & 0 & \frac{3}{4}\mathbf{I}_2 & 0 \\ 0 & 0 & 0 & \frac{3}{4}\mathbf{I}_3 \end{pmatrix} \tag{3.3}$$

There are fairly obvious differences between (3.2), (3.3) and the $\mathbf{Z}$'s and $\mathbf{D}$'s of Section 2. The $\mathbf{Z}$ in (3.2) consists only of the $\mathbf{Z}_{lmm}$ matrices, representing the passage of genes to males from their fathers. The first row of the $\mathbf{Z}$ in Section 2 is lost since there are no measurements of the animals in generation O. The $\mathbf{D}$ matrix in this section has coefficients $\frac{3}{4}$ rather than $\frac{1}{2}$ as in (2.12), because the term $(1 - \frac{1}{2}r_{WW} - \frac{1}{2}r_{XX})(= \frac{1}{2}$, if $r_{WW} = r_{XX} = \frac{1}{2})$ has been replaced by the term $(1 - \frac{1}{2}r_{WW})(= \frac{3}{4}$, if $r_{WW} = \frac{1}{2})$, from (3.1).

The variance matrix $\mathbf{V}$ of $(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3)'$ is now

$$\mathbf{V} = (\mathbf{Z}\Gamma\mathbf{Z}' + \mathbf{I})\sigma^2_e, \tag{3.4}$$

where

$$\Gamma = \mathbf{D}(\sigma^2_A/\sigma^2_e).$$

We now assume a linear model of the form

$$\mathbf{y}_1 = \mathbf{X}_1\alpha + \mathbf{e}_1,$$
$$\mathbf{y}_2 = \mathbf{X}_2\alpha + \mathbf{e}_2,$$
$$\mathbf{y}_3 = \mathbf{X}_3\alpha + \mathbf{e}_3,$$

or more compactly

$$\mathbf{y} = \mathbf{X}\alpha + \mathbf{e}, \tag{3.5}$$

where $\mathbf{y}_i$ is a $N_i \times 1$ vector of records on the *i th* generation, $(1 \leq i \leq t)$, α is a vector representing the unknown fixed effects and $\mathbf{e}$ is a random variable normally distributed with mean zero and variance given by (3.4).

The model is of the same form as that discussed in Sections 2 and 3 of the companion paper (Thompson 1977). Compare (3.4) and (3.5) of this paper with (2.7) of the previous paper, with some changes in notation. We replace γ_A by $\gamma = \sigma^2_A/\sigma^2_e$ and $\mathbf{D}_A$, $\mathbf{w}$ and σ^2 by $\mathbf{D}$, $\mathbf{y}$ and σ^2_e respectively. With these minor changes in notation, the equations in Section 3 of the previous paper could be used to estimate γ, σ^2_e and α. For example, the estimates for γ and σ^2_e compared with (3.7) and (3.8) of the previous paper, found by maximizing the log-likelihood L' of error contrasts (Patterson and Thompson 1971) satisfy

$$\frac{\partial L'}{\partial \gamma} = -(\tfrac{1}{2})\,\mathrm{tr}\,(\mathbf{UD}) + (\tfrac{1}{2})(\beta'\mathbf{D}^{-1}\beta)(\gamma^2\sigma^2_e)^{-1} = 0, \tag{3.6}$$

and

$$\frac{\partial L'}{\partial \sigma^2_e} = -(\tfrac{1}{2})(n-t)/\sigma^2_e + (\mathbf{y}'\mathbf{S}\mathbf{y} - \mathbf{y}'\mathbf{S}\mathbf{Z}\beta)/2\sigma^4_e = 0 \tag{3.7}$$

where

$$\mathbf{S} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}',$$
$$\beta = \mathbf{W}^{-1}\mathbf{Z}'\mathbf{S}\mathbf{y},$$
$$\mathbf{U} = \gamma^{-1}\mathbf{D}^{-1} - \gamma^{-2}\mathbf{D}^{-1}\mathbf{W}^{-1}\mathbf{D}^{-1},$$
$$\mathbf{W} = \mathbf{Z}'\mathbf{S}\mathbf{Z} + \gamma^{-1}\mathbf{D}^{-1},$$

and n and t are the number of observations and the number of fixed effects (for simplicity we assume $\mathbf{X}$ is of full rank). These results follow from using $\Gamma = \gamma\mathbf{D}$.

The fixed effects, α, can be estimated by weighted least squares once γ and σ^2_e are estimated. Once again there is a relationship between α and β that can simplify the computation. The strategy of Section 3 of Thompson (1977) is again useful especially if the number of fixed effects is small compared with the number of sires. This strategy depends on the form of $\mathbf{Z}'\mathbf{Z} + \Gamma^{-1}$. We now investigate the structure of $\mathbf{Z}'\mathbf{Z} + \Gamma^{-1}$ in an attempt to reduce the amount of computation. Taking $\mathbf{Z}$ from (3.2), then $\mathbf{Z}^{*\prime}_0\mathbf{Z}^*_0$ of $\mathbf{Z}'\mathbf{Z}$ is

$$\mathbf{Z}'_{1mm}\mathbf{Z}_{1mm} + \mathbf{Z}'_{1mm}\mathbf{Z}_{1mm}\mathbf{Z}'_{1mm}\mathbf{Z}_{0mm} + \mathbf{Z}'_{0mm}\mathbf{Z}'_{1mm}\mathbf{Z}'_{2mm}\mathbf{Z}_{2mm}\mathbf{Z}_{1mm}\mathbf{Z}_{0mm}. \tag{3.8}$$

The matrices $\mathbf{Z}'_{0mm}$, $\mathbf{Z}'_{0mm}\mathbf{Z}'_{1mm}$ and $\mathbf{Z}'_{0mm}\mathbf{Z}'_{1mm}\mathbf{Z}'_{2mm}$ represent the relationship of individ-

uals in generation one with their fathers, individuals in generation two with their paternal grandfather, and individuals in generation three with their paternal grandfather's father. Each column has only one non-zero element since each individual has only one father, paternal father, paternal grandfather's father, and so on. The non-zero elements are in fact 2^{-j} where j is the number of $\mathbf{Z}$ matrices involved. Hence the *i th* diagonal element of (3.8) is

$$\frac{1}{4}\,(\text{number of sons of sire } i) + \frac{1}{16}\,(\text{number of grandsons of sire } i)$$

$$+ \frac{1}{64}\,(\text{number of great-grandsons of sire } i).$$

The non-diagonal elements of $\mathbf{Z}_0^{*\prime}\mathbf{Z}_0^*$ are zero since no individual is descended from two sires in generation O. Similarly the other submatrices down the diagonal of $\mathbf{Z}'\mathbf{Z}$, $\mathbf{I}_1 + \mathbf{Z}'_{1mm}\mathbf{Z}_{1mm} + \mathbf{Z}'_{1mm}\mathbf{Z}'_{2mm}\mathbf{Z}_{2mm}\mathbf{Z}_{1mm}$, $\mathbf{I}_2 + \mathbf{Z}'_{2mm}\mathbf{Z}_{2mm}$ and $\mathbf{I}_3$ are diagonal and the sub-Matrices of $\mathbf{Z}'\mathbf{Z}$, $\mathbf{Z}_r^{*\prime}\mathbf{Z}_s^*$, have only one non-zero element in each column. These facts enable $\mathbf{Z}'\mathbf{Z} + \Gamma^{-1}$ to be split up into block diagonal form, grouping together animals associated with the *i th* sire in generation zero.

In some cases this subdivision is sufficient to make the problem solvable. In other cases if $\mathbf{Z}'_t\mathbf{Z}_t + \Gamma^{-1}{}_t$ is still a large matrix, it is useful to make further use of the special form of $\mathbf{Z}'_t\mathbf{Z}_t$ and express the inverse in a convenient tabular form (Section 5).

4. Both Sexes Measured

When observations are made on one sex the variance matrix can be inverted quite simply, and the likelihood equations can be set up and solved. However when observations are made on both sexes it is more difficult to find the full ML solution since the variance matrix is more difficult to invert. The pattern of the variance matrix $\mathbf{V}$ (2.13) suggests an opportunist method and this we outline now.

The structure of the variance matrix $\mathbf{V}$ resembles that generated by auto-regressive processes and this suggests that we make a transformation from $\mathbf{y}_t$ to $\mathbf{y}^*{}_t$ where

$$\mathbf{y}_0 = \mathbf{y}^*{}_0,$$

and

$$\mathbf{y}^*{}_i = \mathbf{y}_i - \mathbf{Z}_i\mathbf{y}_{i-1} \qquad (1 \le i \le t).$$

Thus $\mathbf{y}^*{}_t$ represents deviations of the original observations from the parental observations. The variance matrix $\mathbf{V}^*$ is a symmetric block tridiagonal matrix with block elements

$$\begin{aligned}
\mathbf{V}^*{}_{00} &= \mathbf{D}_0\sigma^2{}_A + \mathbf{I}\sigma^2{}_e, \\
\mathbf{V}^*{}_{ii} &= \mathbf{D}_i\sigma^2{}_A + (\mathbf{I} + \mathbf{Z}_i\mathbf{Z}'_i)\sigma^2{}_e, && 1 \le i \le t, \\
\mathbf{V}^*{}_{ij} &= -\mathbf{Z}_i\sigma^2{}_e, && 1 \le i \le t-1, j = i+1, \\
\mathbf{V}^*{}_{ij} &= \mathbf{0}, && 0 \le i,j \le t, |i-j| > 1.
\end{aligned}$$

For example the variance matrix of $\mathbf{y}_0^*, \mathbf{y}_1^*, \mathbf{y}_2^*$ is

$$\mathbf{V}^* = \begin{bmatrix}
\mathbf{D}_0\sigma^2{}_A + \mathbf{I}\sigma^2{}_e & -\mathbf{Z}'_1\sigma^2{}_e & \mathbf{0} \\
-\mathbf{Z}_1\sigma^2{}_e & \mathbf{D}_1\sigma^2{}_A + (\mathbf{I} + \mathbf{Z}_1\mathbf{Z}'_1)\sigma^2{}_e & -\mathbf{Z}'_2\sigma^2{}_e \\
\mathbf{0} & -\mathbf{Z}_2\sigma^2{}_e & \mathbf{D}_2\sigma^2{}_A + (\mathbf{I} + \mathbf{Z}'_2\mathbf{Z}_2)\sigma^2{}_e
\end{bmatrix}$$

When each dam is mated to only one sire, the structure of the variance matrix of $\mathbf{y}^*_t$ is similar to that in a hierarchical analysis of variance with sire and dam components σ^2_e and variance within full-sibs $\sigma^2_e + \sigma^2_A/2$. The covariance between a parental deviation and an offspring deviation is $-\sigma^2_e/2$, and the covariance between deviations two or more generations apart is zero. One possibility is to get estimates of heritability from a hierarchical analysis of variance on y^*_t or from the covariance of y^*_t with y^*_{t+1}. These estimates are not independent and might be combined using arguments similar to those of Hill and Nicholas (1974), who consider the combination of parent-offspring and sib-covariance estimates of heritability.

5. Comments

The revision of the paper contained an appendix, reviewed by the referees and now available from the author. In this appendix we show how the inverse of the submatrix of $\mathbf{Z'Z} + \Gamma^{-1}$ associated with the *ith* sire in generation zero can be written as the product of an upper triangular matrix, $\mathbf{F}$, and its transpose. We also give a numerical example to illustrate the calculations and present a tabular method for the formation of $\mathbf{F}$.

Acknowledgments

I am very grateful to Oscar Kempthorne and a referee for helpful comments.

L'ESTIMATION DE L'HERITABILITE A PARTIR DE DONNEES NON EQUILIBREES

II—Données relevées sur plus de deux générations

Résumé

On considère l'estimation de l'héritabilité par le maximum de vraisemblance dans le cas où les données sont recueillies sur un seul sexe et pour plus de deux générations, et que les descendants sont des demi-frères (ou soeurs) sans consanguinité. On obtient les covariances entre apparentés dans différentes générations. Les équations du maximum de vraisemblance sont données en utilisant les résultats obtenus dans un précédent article. Quelques unes des formules obtenues dépendent de l'inversion de matrices et on donne une présentation tabulée appropriée pour obtenir ces matrices inverses en tirant parti de leur structure spéciale. On discute brièrement de l'estimation dans le cas où les données sont recueillies sur les deux sexes.

References

Graybill, F. A. (1969). *Introduction to Matrices with Applications in statistics,* Wadsworth, California.

Henderson, C. R. (1976). A simple method for computing the inverse of a numerator matrix used in prediction of breeding values, *Biometrics 32,* 69–84.

Hill, W. G. (1971). Design and efficiency of selection experiments for estimating genetic parameters, *Biometrics 27,* 293–311.

Hill, W. G. (1974). Prediction and evaluation of response to selection with overlapping generations, *Animal Production 18,* 117–139.

Hill, W. G. and Nicholas, F. W. (1974). Estimation of heritability by both regression of offspring on parent and intra-class correlation of sibs in one experiment, *Biometrics 30,* 447–468.

Kempthorne, O. (1957). *An Introduction to Genetic Statistics.* Wiley, New York.

Malécot, G. (1948). *Les Mathematiques de l'Héredité.* Masson et Cie, Paris.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal, *Biometrika 58,* 545–554.

Thompson, R. (1977). The estimation of heritability with unbalanced data. I. Observations available on parent and offspring, *Biometrics 33,* page nos. 485–495.

Received March 1976; Revised October 1976

Reprinted from
BIOMETRICS COPYRIGHT © 1977
THE BIOMETRIC SOCIETY, Vol. 33, No. 3, September 1977

20

The estimation of heritability with unbalanced data. III. Unpublished Appendices, 1-17

by

R. Thompson

The Estimation of Heritability with Unbalanced Data

III Unpublished Appendices

Robin Thompson

ARC Unit of Statistics, University of Edinburgh

Edinburgh EH9 3JZ, Scotland

Summary

Appendices submitted with parts I and II of this study were not published due to pressure of space. They are given and also FORTRAN coding to do the calculations mentioned in II.

APPENDIX A - Proof of (3.11)-(3.13) in I.

We now prove (3.11)-(3.13). By using (A4) we can write $\mathrm{tr}(\underset{\sim}{U}\underset{\sim}{D}_j)$ as

$$\mathrm{tr}(\underset{\sim}{U}^*\underset{\sim}{D}_j) - \mathrm{tr}(\underset{\sim}{\Gamma}^{-1}(\underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1})^{-1}\underset{\sim}{Z}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{H}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Z}(\underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1})^{-1}\underset{\sim}{\Gamma}^{-1}\underset{\sim}{D}_j)$$

$$= e_j - \mathrm{tr}((\underset{\sim}{X}'\underset{\sim}{H}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Z}(\underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1})^{-1}\underset{\sim}{\Gamma}^{-1}\underset{\sim}{D}_j\underset{\sim}{\Gamma}^{-1}(\underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1})^{-1}\underset{\sim}{Z}'\underset{\sim}{X})$$

$$= e_j - \mathrm{tr}(\underset{\sim}{Q}\,(\underset{\sim}{E}_j)_{11}) \quad \text{(i.e. (3.11))}.$$

Now $(\underset{\sim}{w} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}}) = (\underset{\sim}{X}\ \ \underset{\sim}{w})\begin{pmatrix} -\hat{\underset{\sim}{\alpha}} \\ 1 \end{pmatrix} = \underset{\sim}{Y}\underset{\sim}{\alpha}_*$ (using (A6)),

where $$\underset{\sim}{\alpha}_*' = (-\hat{\underset{\sim}{\alpha}}'\ \ 1)\ . \qquad \text{(A15)}$$

From (A2) $\tilde{\underset{\sim}{\beta}}'\underset{\sim}{\Gamma}^{-1}\underset{\sim}{D}_j\underset{\sim}{\Gamma}^{-1}\tilde{\underset{\sim}{\beta}}$ can be written as

$$(\underset{\sim}{w} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}})'\underset{\sim}{Z}(\underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1})^{-1}\underset{\sim}{\Gamma}^{-1}\underset{\sim}{D}_j\underset{\sim}{\Gamma}^{-1}(\underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1})^{-1}\underset{\sim}{Z}'(\underset{\sim}{w} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}})$$

$$= \underset{\sim}{\alpha}_*'\underset{\sim}{Y}\underset{\sim}{Z}(\underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1})^{-1}\underset{\sim}{\Gamma}^{-1}\underset{\sim}{D}_j\underset{\sim}{\Gamma}^{-1}(\underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1})^{-1}\underset{\sim}{Z}'\underset{\sim}{Y}'\underset{\sim}{\alpha}_*$$

$$= \underset{\sim}{\alpha}_*' \underset{\sim}{E}_j \underset{\sim}{\alpha}_* \quad \text{(i.e. (3.12))}.$$

Also $$\underset{\sim}{w}'\underset{\sim}{S}\underset{\sim}{w} - \underset{\sim}{w}'\underset{\sim}{S}\underset{\sim}{Z}\tilde{\underset{\sim}{\beta}} = \underset{\sim}{w}'\underset{\sim}{w} - \underset{\sim}{w}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{w} - \underset{\sim}{w}'\underset{\sim}{S}\underset{\sim}{Z}\tilde{\underset{\sim}{\beta}}$$

$$= \underset{\sim}{w}'\underset{\sim}{w} - \underset{\sim}{w}'\underset{\sim}{X}\hat{\underset{\sim}{\alpha}} - \underset{\sim}{w}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Z}\tilde{\underset{\sim}{\beta}} - \underset{\sim}{w}'\underset{\sim}{Z}\tilde{\underset{\sim}{\beta}} + \underset{\sim}{w}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Z}\tilde{\underset{\sim}{\beta}} \quad \text{(using (A1) and (3.3))}$$

$$= \underset{\sim}{w}'\underset{\sim}{w} - \underset{\sim}{w}'\underset{\sim}{X}\hat{\underset{\sim}{\alpha}} - \underset{\sim}{w}'\underset{\sim}{Z}(\underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1})^{-1}\underset{\sim}{Z}'\underset{\sim}{w} + \underset{\sim}{w}'\underset{\sim}{Z}(\underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1})^{-1}\underset{\sim}{Z}'\underset{\sim}{X}\hat{\underset{\sim}{\alpha}}$$

$$= (\underset{\sim}{B})_{22} - (\underset{\sim}{B})_{21}\hat{\underset{\sim}{\alpha}} - (\underset{\sim}{C})_{22} + (\underset{\sim}{C})_{21}\hat{\underset{\sim}{\alpha}} \quad \text{(i.e. (3.13))}.$$

Terms involved in the second differentials ((3.17) and (3.18)) can be derived in a similar way.

APPENDIX B - Extension when $\underset{\sim}{X}$ and $\underset{\sim}{Z}$ contain fixed and random effects

We now show how the results of section 3 can be modified to deal with the case mentioned in section 4, when $\underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1}$ can be split up into block diagonal form, and $\underset{\sim}{X}$ and $\underset{\sim}{Z}$ contain both fixed and random effects. Suppose we have a model of the form

$$\underset{\sim}{w} = \underset{\sim f}{X}\underset{\sim}{\alpha}_f + \underset{\sim f}{Z}\underset{\sim}{\beta}_f + \underset{\sim}{e}$$

with $\underset{\sim}{e}$ having normal distribution with variance matrix $\underset{\sim}{V} = (\underset{\sim r}{X}'\underset{\sim}{\Gamma}\underset{\sim x}{X}\underset{\sim r}{} + \underset{\sim r}{Z}'\underset{\sim}{\Gamma}\underset{\sim z}{Z}\underset{\sim r}{} + \underset{\sim}{I})\sigma^2$. We let $\underset{\sim}{X} = (\underset{\sim r}{X} \quad \underset{\sim f}{X})$, $\underset{\sim}{Z} = (\underset{\sim r}{Z} \quad \underset{\sim f}{Z})$

$$\underset{\sim}{\Gamma} = \begin{pmatrix} \underset{\sim z}{\Gamma} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix} , \quad \underset{\sim}{\Gamma}^{-1} = \begin{pmatrix} \underset{\sim z}{\Gamma}^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix} ,$$

$$\underset{\sim 1}{\Gamma} = \begin{pmatrix} \underset{\sim x}{\Gamma} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix} , \quad \underset{\sim 1}{\Gamma}^{-1} = \begin{pmatrix} \underset{\sim x}{\Gamma}^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix} ,$$

$\underset{\sim}{\Gamma} = \sum_{\ell} \underset{\sim z\ell}{D}\gamma_{z\ell}$ and $(\underset{\sim 1}{\Gamma})_{11} = \sum_{\ell} \underset{\sim x\ell}{D}\gamma_{x\ell}$ (using the notation of (A13)), where $\underset{\sim}{\Gamma}$ and $\underset{\sim 1}{\Gamma}$ are $b \times b$ and $(t+1) \times (t+1)$ matrices respectively, and $\gamma_{x\ell}$ and $\gamma_{z\ell}$ are unknown variance parameters. If $\underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1}$ is block diagonal, then we can again use the results of section 3 with minor modifications. We add $\underset{\sim 1}{\Gamma}^{-1}$ to $\underset{\sim}{B}$ in (A7). Again (3.10)-(3.13), (3.17) and (3.18) are relevant although $\hat{\underset{\sim}{\alpha}}$ now contains predictors of random effects as well as estimates of fixed effects. We can use (3.10)-(3.13), (3.7) and (3.18) to give the relevant terms in the likelihood equations relating to the random effects in the $\underset{\sim}{Z}$ part of the model. The corresponding terms for the random effects in the $\underset{\sim}{X}$ part can be found from (3.17), (3.14) and (3.15) by replacing $\underset{\sim}{U}$ by $(\underset{\sim 1}{\Gamma}^{-1})_{11} - (\underset{\sim 1}{\Gamma}^{-1})_{11}\underset{\sim}{Q}(\underset{\sim 1}{\Gamma}^{-1})_{11}$, $\underset{\sim}{\Gamma}^{-1}$ by $(\underset{\sim 1}{\Gamma}^{-1})_{11}$ and $\tilde{\underset{\sim}{\beta}}$ by $\underset{\sim}{\alpha}$.

Terms similar to $\mathrm{tr}(\underset{\sim}{U}\underset{\sim j}{D}\underset{\sim}{U}\underset{\sim \ell}{D})$ are required to evaluate the second differential with respect to γ_{xj} and $\gamma_{z\ell}$ and we now show they are of

the form $\mathrm{tr}((E_\ell)_{11}\underset{\sim}{Q}(\Gamma_1^{-1})_{11}D_{Xj}(\Gamma_1^{-1})_{11}\underset{\sim}{Q})$. We let (by analogy with (3.5) and (3.6))

$$\begin{pmatrix} U_{XX} & U_{XZ} \\ U_{ZX} & U_{ZZ} \end{pmatrix} = \begin{pmatrix} (\Gamma_1)_{11} & 0 \\ 0 & \Gamma \end{pmatrix}^{-1} - \begin{pmatrix} (\Gamma_1)_{11} & 0 \\ 0 & \Gamma \end{pmatrix}^{-1} W \begin{pmatrix} (\Gamma_1)_{11} & 0 \\ 0 & \Gamma \end{pmatrix}^{-1}$$

and $$W = \begin{pmatrix} X_r'X_r + \Gamma_x^{-1} & X_r'X_f & X_r'Z_r & X_r'Z_f \\ X_f'X_r & X_f'X_f & X_f'Z_r & X_f'Z_f \\ Z_r'X_r & Z_r'X_f & Z_r'Z_r + \Gamma_z^{-1} & Z_r'Z_f \\ Z_f'X_r & Z_f'X_f & Z_f'Z_r & Z_f'Z_f \end{pmatrix}$$

We require (compare with (3.14))

$$\mathrm{tr}\left[\begin{pmatrix} U_{XX} & U_{XZ} \\ U_{ZX} & U_{ZZ} \end{pmatrix}\begin{pmatrix} D_{Xj} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} U_{XX} & U_{XZ} \\ U_{ZX} & U_{ZZ} \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & D_{Z\ell} \end{pmatrix}\right]$$

$$= \mathrm{tr}\left[\begin{pmatrix} U_{XX}D_{Xj} & 0 \\ U_{ZX}D_{Xj} & 0 \end{pmatrix}\begin{pmatrix} 0 & U_{XZ}D_{Z\ell} \\ 0 & U_{ZZ}D_{Z\ell} \end{pmatrix}\right] = \mathrm{tr}\begin{pmatrix} 0 & U_{XX}D_{Xj}U_{XZ}D_{Z\ell} \\ 0 & U_{ZX}D_{Xj}U_{XZ}D_{Z\ell} \end{pmatrix} = \mathrm{tr}\left[U_{ZX}D_{Xj}U_{XZ}D_{Z\ell}\right] .$$

Since the inverse of W is of the form

$$W^{-1} = \begin{pmatrix} Q & -Q'X'Z(Z'Z + \Gamma^{-1})^{-1} \\ -(Z'Z + \Gamma^{-1})^{-1}Z'XQ & \cdots \end{pmatrix}$$

$\mathrm{tr}\left[U_{ZX}D_{Xj}U_{XZ}D_{Z\ell}\right]$ can be expanded as

$$\mathrm{tr}\left[\Gamma^{-1}(Z'Z + \Gamma^{-1})^{-1}Z'XQ(\Gamma_1^{-1})_{11}D_{Xj}(\Gamma_1^{-1})_{11}QX'Z(Z'Z + \Gamma^{-1})^{-1}(\Gamma_1^{-1})_{11}D_{Z\ell}\right]$$

$$= \mathrm{tr}\left[X'Z(Z'Z + \Gamma^{-1})^{-1}\Gamma^{-1}D_{Z\ell}\Gamma^{-1}(Z'Z + \Gamma^{-1})^{-1}Z'XQ(\Gamma_1^{-1})_{11}D_{Xj}(\Gamma_1^{-1})_{11}Q\right]$$

$$= \mathrm{tr}\left[(E_\ell)_{11}Q(\Gamma_1^{-1})_{11}D_{Xj}(\Gamma_1^{-1})_{11}Q\right]$$

and so is in the stated form.

APPENDIX C - (From II)

In this appendix we show how the inverse of the sub-matrix of $\underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1}$ associated with the i-th sire in generation 0 can be written as the product of an upper triangular matrix, $\underset{\sim}{F}$, and its transpose. We also give a numerical example to illustrate the calculations and present a tabular method for the formation of $\underset{\sim}{F}$. Let $\underset{\sim}{C}$ be the symmetric sub-matrix of $\underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1}$ associated with the i-th sire in generation 0. We can partition $\underset{\sim}{C}$ into a matrix with terms association with each generation of the form

$$\begin{pmatrix} \underset{\sim}{C}_{00} & \underset{\sim}{C}_{01} & \underset{\sim}{C}_{02} & \underset{\sim}{C}_{03} \\ \underset{\sim}{C}_{10} & \underset{\sim}{C}_{11} & \underset{\sim}{C}_{12} & \underset{\sim}{C}_{13} \\ \underset{\sim}{C}_{20} & \underset{\sim}{C}_{21} & \underset{\sim}{C}_{22} & \underset{\sim}{C}_{23} \\ \underset{\sim}{C}_{30} & \underset{\sim}{C}_{31} & \underset{\sim}{C}_{32} & \underset{\sim}{C}_{33} \end{pmatrix} . \qquad \text{(A1)}$$

We require the inverse of $\underset{\sim}{C}$. One possibility is to write $\underset{\sim}{C}$ as the product of a triangular matrix and its transpose. The Choleski method is sometimes used to express a symmetric matrix as the product of a lower triangular matrix and its transpose (for instance, Graybill [1969] p.298). We require a slight extension of this method to deal with sub-matrices, and also find that to take advantage of the structure it is easier to express $\underset{\sim}{C}$ as the product of an upper triangular matrix and its transpose.

If

$$\underset{\sim}{C} = \underset{\sim}{T}\underset{\sim}{T}'$$

the sub-matrices of $\underset{\sim}{T}$ are given recursively by (if there are t generations)

$$\underset{\sim}{T}_{tt} = (\underset{\sim}{C}_{tt})^{\frac{1}{2}} , \qquad \text{(A2)}$$

$$\underset{\sim}{T}_{rt} = \underset{\sim}{C}_{rt}(T_{tt})^{-1} , \qquad 0 \leq r \leq t-1 , \quad \text{(A3)}$$

$$\underset{\sim}{T}_{rr} = (\underset{\sim}{C}_{rr} - \sum_{q=r+1}^{t} \underset{\sim}{T}_{rq}\underset{\sim}{T}'_{rq})^{\frac{1}{2}} , \quad \text{(A4)}$$

$$\underset{\sim}{T}_{sr} = (\underset{\sim}{C}_{sr} - \sum_{q=r+1}^{t} \underset{\sim}{T}_{sq}\underset{\sim}{T}'_{rq})(T_{rr})^{-1} , \quad 0 \leq s \leq r-1, \quad \text{(A5)}$$

where (A4) is evaluated for $r = t - 1$, (A5) for $0 \leq s \leq t - 2$, then (A4) for $r = t - 2$, (A5) for $0 \leq s \leq t - 3$, etc. Because of the structure of the $\underset{\sim}{C}$ matrix, $\underset{\sim}{C}_{tt}$ and $\underset{\sim}{C}_{rr} - \sum_{q=r+1}^{t} \underset{\sim}{T}_{rq}\underset{\sim}{T}'_{rq}$ are diagonal matrices so that the operation of taking the square root, i.e. $(\underset{\sim}{C}_{tt})^{\frac{1}{2}}$, is equivalent to taking the square root of the diagonal elements.

The inverse of $\underset{\sim}{T}$ is an upper triangular matrix $\underset{\sim}{F}$, where $\underset{\sim}{F}$ can be partitioned and found recursively as follows:

$$\underset{\sim}{F}_{rr} = \underset{\sim}{T}_{rr}^{-1} , \quad \text{(A6)}$$

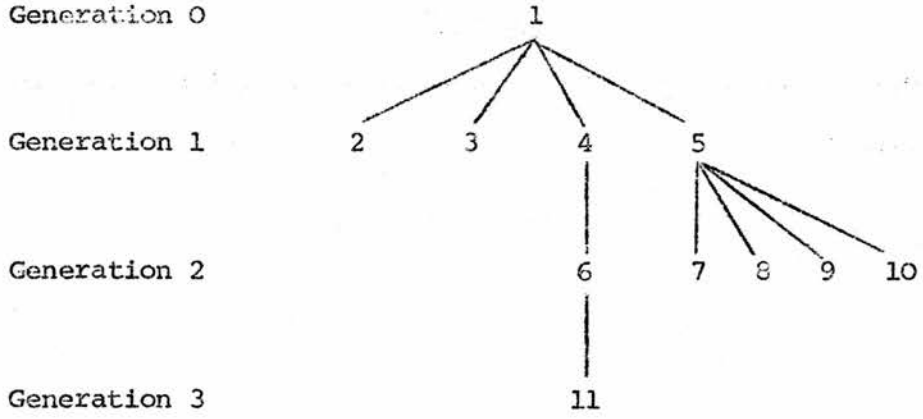
$$\underset{\sim}{F}_{rs} = -\left(\sum_{q=r}^{s-1} \underset{\sim}{F}_{rq}\underset{\sim}{T}_{qs}\right)\underset{\sim}{F}_{ss} , \qquad r < s \leq t , \quad \text{(A7)}$$

where (A6) is evaluated for $r = t, t - 1, \ldots, 0$ and (A7) is evaluated for $r = t - 1$, $t - 1 \leq s \leq t$, $r = t - 2$, $t - 2 \leq s \leq t$ etc.

Now $\underset{\sim}{C}^{-1}$ is of the form $\underset{\sim}{F}\underset{\sim}{F}'$ and so the results of section 3 of Thompson [1977] can be used.

FIGURE 1

Pedigree of eleven individuals



To illustrate the calculations involved in equations (A2)-(A7) we give an example based on the pedigree in Figure 1. Measurements are available on individuals numbered 2-11. Individuals 2-5 are sons of sire 1, 6 is the son of 4, 7-10 are sons of 5, and 11 is the son of 6. There are respectively 1, 4, 5 and 1 individuals in the 0-th, 1st, 2nd and 3rd generations so, $\underset{\sim}{Z}_{0mm}$, $\underset{\sim}{Z}_{1mm}$ and $\underset{\sim}{Z}_{2mm}$ have sizes 4×1, 5×4 and 1×5. Now

$$\underset{\sim}{Z}_{0mm} = (\tfrac{1}{2} \quad \tfrac{1}{2} \quad \tfrac{1}{2} \quad \tfrac{1}{2})' ,$$

$$\underset{\sim}{Z}_{1mm} = \begin{pmatrix} 0 & 0 & \tfrac{1}{2} & 0 \\ 0 & 0 & 0 & \tfrac{1}{2} \\ 0 & 0 & 0 & \tfrac{1}{2} \\ 0 & 0 & 0 & \tfrac{1}{2} \\ 0 & 0 & 0 & \tfrac{1}{2} \end{pmatrix} , \quad \underset{\sim}{Z}_{1mm}\underset{\sim}{Z}_{0mm} = (\tfrac{1}{8} \quad \tfrac{1}{4} \quad \tfrac{1}{4} \quad \tfrac{1}{4} \quad \tfrac{1}{4})' ,$$

$$\underset{\sim}{Z}_{2mm} = (\tfrac{1}{2} \quad 0 \quad 0 \quad 0 \quad 0) , \quad \underset{\sim}{Z}_{2mm}\underset{\sim}{Z}_{1mm} = (0 \quad 0 \quad \tfrac{1}{4} \quad 0) ,$$

$$\underset{\sim}{Z}_{2mm}\underset{\sim}{Z}_{1mm}\underset{\sim}{Z}_{0mm} = (\tfrac{1}{8}) ,$$

so that $\underset{\sim}{Z}$ of (3.2) is

$$\underset{\sim}{Z} = \begin{pmatrix} \tfrac{1}{2} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \tfrac{1}{2} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \tfrac{1}{2} & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \tfrac{1}{2} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \tfrac{1}{4} & 0 & 0 & \tfrac{1}{2} & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \tfrac{1}{4} & 0 & 0 & 0 & \tfrac{1}{2} & 0 & 1 & 0 & 0 & 0 & 0 \\ \tfrac{1}{4} & 0 & 0 & 0 & \tfrac{1}{2} & 0 & 0 & 1 & 0 & 0 & 0 \\ \tfrac{1}{4} & 0 & 0 & 0 & \tfrac{1}{2} & 0 & 0 & 0 & 1 & 0 & 0 \\ \tfrac{1}{4} & 0 & 0 & 0 & \tfrac{1}{2} & 0 & 0 & 0 & 0 & 1 & 0 \\ \tfrac{1}{8} & 0 & 0 & \tfrac{1}{4} & 0 & \tfrac{1}{2} & 0 & 0 & 0 & 0 & 1 \end{pmatrix} .$$

We usually require the inverse of $\underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1}$, where $\underset{\sim}{\Gamma}^{-1}$ is a diagonal matrix that is a function of γ. However, for computational simplicity in showing the form of equations (A2)-(A7), we let Γ^{-1} be a matrix with all elements zero except the first element which we take to be 64, and evaluate $64\underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1}$.

Forming $64\underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1}$ and partitioning the resultant matrix as in (A1) we find

$$\underset{\sim}{C}_{00} = (149) = (64(4/4 + 5/16 + 1/64) + 64),$$

$$\underset{\sim}{C}_{01} = (32 \quad 32 \quad 42 \quad 64),$$

$$\underset{\sim}{C}_{02} = (20 \quad 16 \quad 16 \quad 16 \quad 16),$$

$$\underset{\sim}{C}_{03} = (8),$$

$$\underset{\sim}{C}_{11} = \begin{pmatrix} 64 & 0 & 0 & 0 \\ 0 & 64 & 0 & 0 \\ 0 & 0 & 84 & 0 \\ 0 & 0 & 0 & 128 \end{pmatrix},$$

$$\underset{\sim}{C}_{12} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 40 & 0 & 0 & 0 & 0 \\ 0 & 32 & 32 & 32 & 32 \end{pmatrix}, \quad \underset{\sim}{C}_{13} = (0 \quad 0 \quad 16 \quad 0)',$$

$$\underset{\sim}{C}_{22} = \begin{pmatrix} 80 & 0 & 0 & 0 & 0 \\ 0 & 64 & 0 & 0 & 0 \\ 0 & 0 & 64 & 0 & 0 \\ 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 64 \end{pmatrix}, \quad \underset{\sim}{C}_{23} = (32 \quad 0 \quad 0 \quad 0 \quad 0)',$$

and $\underset{\sim}{C}_{33} = (64)$.

We now find $\underset{\sim}{T}$ such that $\underset{\sim}{C} = \underset{\sim}{T}\underset{\sim}{T}'$.

The sub-matrices of $\underset{\sim}{T}$ satisfy (using (A2)-(A5))

$$\underset{\sim}{T}_{33} = (\underset{\sim}{C}_{33})^{\frac{1}{2}} = (8), \quad \underset{\sim}{T}_{23} = \underset{\sim}{C}_{23}\underset{\sim}{T}_{33}^{-1} = (4 \quad 0 \quad 0 \quad 0 \quad 0)',$$

$$\underset{\sim}{T}_{13} = \underset{\sim}{C}_{13}\underset{\sim}{T}_{33}^{-1} = (0 \quad 0 \quad 2 \quad 0)', \quad \underset{\sim}{T}_{03} = \underset{\sim}{C}_{03}\underset{\sim}{T}_{33}^{-1} = (1),$$

$$\underset{\sim}{T}_{22} = (\underset{\sim}{C}_{22} - \underset{\sim}{T}_{23}\underset{\sim}{T}_{23}')^{\frac{1}{2}} = 8\underset{\sim}{I},$$

$$\underline{T}_{12} = (\underline{C}_{12} - \underline{T}_{13}\underline{T}'_{23})\underline{T}_{22}^{-1} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 \\ 0 & 4 & 4 & 4 & 4 \end{pmatrix},$$

$$\underline{T}_{02} = (\underline{C}_{02} - \underline{T}_{03}\underline{T}'_{23})\underline{T}_{22}^{-1} = (2 \quad 2 \quad 2 \quad 2 \quad 2),$$

$$\underline{T}_{11} = (\underline{C}_{11} - \underline{T}_{12}\underline{T}'_{21} - \underline{T}_{13}\underline{T}'_{31})^{\frac{1}{2}} = 8\underline{I},$$

$$\underline{T}_{01} = (\underline{C}_{01} - \underline{T}_{02}\underline{T}'_{12} - \underline{T}_{03}\underline{T}'_{13})\underline{T}_{11}^{-1} = (4 \quad 4 \quad 4 \quad 4),$$

$$\underline{T}_{00} = (\underline{C}_{00} - \underline{T}_{01}\underline{T}'_{01} - \underline{T}_{02}\underline{T}'_{02} - \underline{T}_{03}\underline{T}'_{03})^{\frac{1}{2}} = 8\underline{I}.$$

Equations (A6) and (A7) can be used now to find $\underline{F}$, the inverse of $\underline{T}$, again in partitioned form.

$$\underline{F}_{33} = \underline{T}_{33}^{-1} = (1/8)\underline{I}, \quad \underline{F}_{22} = \underline{T}_{22}^{-1} = (1/8)\underline{I},$$

$$\underline{F}_{23} = -\underline{F}_{22}\underline{T}_{23}\underline{F}_{33} = (-1/16)(1 \quad 0 \quad 0 \quad 0 \quad 0)',$$

$$\underline{F}_{11} = \underline{T}_{11}^{-1} = (1/8)\underline{I},$$

$$\underline{F}_{12} = -\underline{F}_{11}\underline{T}_{12}\underline{F}_{22} = (-1/16)\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \end{pmatrix},$$

$$\underline{F}_{13} = -(\underline{F}_{11}\underline{T}_{13} + \underline{F}_{12}\underline{T}_{23})\underline{F}_{33} = (0 \quad 0 \quad 0 \quad 0)',$$

$$\underline{F}_{00} = \underline{T}_{00}^{-1} = (1/8)\underline{I},$$

$$\underline{F}_{01} = -\underline{F}_{00}\underline{T}_{01}\underline{F}_{11} = (1/16)(1 \quad 1 \quad 1 \quad 1),$$

$$\underline{F}_{02} = -(\underline{F}_{00}\underline{T}_{02} + \underline{F}_{01}\underline{T}_{12})\underline{F}_{22} = (0 \quad 0 \quad 0 \quad 0 \quad 0),$$

$$\underline{F}_{03} = -(\underline{F}_{00}\underline{T}_{03} + \underline{F}_{01}\underline{T}_{13} + \underline{F}_{02}\underline{T}_{23})\underline{F}_{33} = (0),$$

so that

$$16\underline{F} = \begin{bmatrix} 2 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & -1 & -1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix} . \qquad (A8)$$

Since a large number of the elements of the matrices are zero we present the results in a convenient tabular form that is relatively easy to program. We first define an incidence matrix $\underline{N}$ of size $(t + 1) \times n$, where $t + 1$ is the number of generations involved and n the number of individuals. The j-th column represents the pedigree of the j-th individual and the i-th row contains individuals in the $(i - 1)$-th generation. For the numerical example $t = 3$ and $n = 11$ and

$$\underline{N} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 & 5 & 4 & 5 & 5 & 5 & 5 & 4 \\ 0 & 0 & 0 & 0 & 0 & 6 & 7 & 8 & 9 & 10 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 11 \end{bmatrix} . \qquad (A9)$$

For convenience we fill up the undefined elements of $\underline{N}$ with zeroes. We now form a reduced matrix $R(\underline{Z}'\underline{Z})$, again of size $(t + 1) \times n$, from the upper triangular part of $\underline{Z}'\underline{Z}$ such that $R(\underline{Z}'\underline{Z})_{ij} = (\underline{Z}'\underline{Z})_{N_{ij},j}$.

In the numerical example

$$R(64\underline{Z}'\underline{Z}) = \begin{pmatrix} 85 & 32 & 32 & 42 & 64 & 20 & 16 & 16 & 16 & 16 & 8 \\ 0 & 64 & 64 & 84 & 128 & 40 & 32 & 32 & 32 & 32 & 16 \\ 0 & 0 & 0 & 0 & 0 & 80 & 64 & 64 & 64 & 64 & 32 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 64 \end{pmatrix}, \quad \text{(A10)}$$

where, for example, 40, the (2,6) element of $R(64\underline{Z}'\underline{Z})$, is the (4,6) element of $64\underline{Z}'\underline{Z}$ where 4 is the (2,6) element of $\underline{N}$. Note that the reduced matrix contains all the non-zero elements of $64\underline{Z}'\underline{Z}$ and that again we fill up undefined elements with zeroes. In a similar way reduced matrices $R(\underline{C})$, $R(\underline{T})$ and $R(\underline{F})$ of the upper triangular part of $\underline{C}$ and the upper triangular matrices $\underline{T}$ and $\underline{F}$ can be defined. For the numerical example we find

$$R(\underline{C}) = \begin{pmatrix} 149 & 32 & 32 & 42 & 64 & 20 & 16 & 16 & 16 & 16 & 8 \\ 0 & 64 & 64 & 84 & 128 & 40 & 32 & 32 & 32 & 32 & 16 \\ 0 & 0 & 0 & 0 & 0 & 80 & 64 & 64 & 64 & 64 & 32 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 64 \end{pmatrix}, \quad \text{(A11)}$$

$$R(\underline{T}) = \begin{pmatrix} 8 & 4 & 4 & 4 & 4 & 2 & 2 & 2 & 2 & 2 & 1 \\ 0 & 8 & 8 & 8 & 8 & 4 & 4 & 4 & 4 & 4 & 2 \\ 0 & 0 & 0 & 0 & 0 & 8 & 8 & 8 & 8 & 8 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 \end{pmatrix}, \quad \text{(A12)}$$

$$R(16\underline{F}) = \begin{pmatrix} 2 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 2 & 2 & 2 & -1 & -1 & -1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 2 & 2 & 2 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}. \quad \text{(A13)}$$

We also let g_j be the number of non-zero elements in the j-th column of $\underset{\sim}{N}$ and note that the j-th individual is in generation $g_j - 1$. The (g_j, j) elements of the reduced matrices represent the j-th diagonal elements of the original matrices. For instance, if $j = 6$, $g_j = 3$ and the (3,6) element of $R(\underset{\sim}{C})$ is the sixth diagonal element of $\underset{\sim}{C}$. A convenient way to form $R(\underset{\sim}{Z}'\underset{\sim}{Z})$ is to form first the diagonal elements of $\underset{\sim}{Z}'\underset{\sim}{Z}$. These elements are given by

$$(R(\underset{\sim}{Z}'\underset{\sim}{Z}))_{g_k,k} = \Sigma' r_j \times 4^{-(g_j - g_k)}, \qquad (k = 1, \ldots, n), \qquad (A14)$$

where Σ' indicates summation over values of j such that $N_{g_k j} = k$ and r_j is 1 if the j-th individual is observed and 0 if it is not. The summation, for the k-th diagonal element, is over those elements of $N_{g_k j}$ for which $N_{g_k j} = k$. In the numerical example $r_1 = 0$, $r_j = 1$ $(j > 1)$ and

$$(R(\underset{\sim}{Z}'\underset{\sim}{Z}))_{1,1} = (\frac{0}{1} + \frac{4}{4} + \frac{5}{16} + \frac{1}{64}) = \frac{85}{64}.$$

The other elements of $R(\underset{\sim}{Z}'\underset{\sim}{Z})$ satisfy

$$(R(\underset{\sim}{Z}'\underset{\sim}{Z}))_{ij} = \tfrac{1}{4}(R(\underset{\sim}{Z}'\underset{\sim}{Z}))_{(i+1),j} \qquad (i < g_j) \qquad (A15)$$

and so can be calculated recursively. Since $\underset{\sim}{C} = \underset{\sim}{Z}'\underset{\sim}{Z} + \underset{\sim}{\Gamma}^{-1}$ and $\underset{\sim}{\Gamma}^{-1}$ is a diagonal matrix $R(\underset{\sim}{C})$ can be calculated by adding the elements of $\underset{\sim}{\Gamma}^{-1}$ to the appropriate elements of $R(\underset{\sim}{Z}'\underset{\sim}{Z})$.

$R(\underset{\sim}{T})$ can be found recursively, in the same order as the block matrices of $\underset{\sim}{T}$ were found, by rewriting equations (A2) to (A5) in terms of $R(\underset{\sim}{T})$ and $R(\underset{\sim}{C})$. They become

$$(R(\underset{\sim}{T}))_{(t+1),j} = [(R(\underset{\sim}{C}))_{(t+1),j}]^{\frac{1}{2}}, \qquad j \text{ s.t. } g_j = t + 1, \qquad (A16)$$

$$(R(\underset{\sim}{T}))_{rj} = ((R(\underset{\sim}{C}))_{rj})/((R(\underset{\sim}{T}))_{(t+1),j}), \qquad j \text{ s.t. } g_j = t + 1 \qquad (A17)$$

$$(R(\underset{\sim}{T}))_{rj} = [(R(\underset{\sim}{C}))_{rj} - \Sigma'' \; (R(\underset{\sim}{T}))_{rq}(R(\underset{\sim}{T}))_{rq}]^{\frac{1}{2}} ,$$

$$j \text{ s.t.} g_j = r \ , \ 1 \leq r \leq t \ , \tag{A18}$$

$$(R(\underset{\sim}{T}))_{sj} = [(R(\underset{\sim}{C}))_{sj} - \Sigma'' \; (R(\underset{\sim}{T}))_{rq}(R(\underset{\sim}{T}))_{sq}]/(R(\underset{\sim}{T}))_{rj} ,$$

$$j \text{ s.t.} g_j = r \ , \ 1 \leq s \leq r-1 \ , \ 1 \leq r \leq t \ , \tag{A19}$$

where j s.t. $g_t = t+1$ indicates that (A8) and (A9) define elements in the columns corresponding to animals in the t-th generation. Similarly (A10) and (A11) define elements in the columns corresponding to animals in the $(r-1)$-th generation. Σ'' indicates summation over values of q such that $N_{rq} = j$ and $g_q > g_j$ and can be interpreted as calculating $(R(\underset{\sim}{T}))_{rq}(R(\underset{\sim}{T}))_{sq}$ for each q such that $g_q > g_j$ and subtracting this product from $(R(\underset{\sim}{C}))_{sj}$ where j is N_{rq}.

For instance, $(R(\underset{\sim}{T}))_{4,11} = [(R(\underset{\sim}{C}))_{4,11}]^{\frac{1}{2}} = [64]^{\frac{1}{2}} = 8$,

$$(R(\underset{\sim}{T}))_{3,6} = [(R(\underset{\sim}{C}))_{3,6} - (R(\underset{\sim}{T}))_{3,10})^2]^{\frac{1}{2}}$$

$$= [80 - 16]^{\frac{1}{2}} = 8 \ ,$$

$$(R(\underset{\sim}{T}))_{2,4} = [(R(\underset{\sim}{C}))_{2,4} - ((R(\underset{\sim}{T}))_{2,6})^2 - ((R(\underset{\sim}{T}))_{2,11})^2]^{\frac{1}{2}}$$

$$= [84 - 16 - 4]^{\frac{1}{2}} = 8 \ ,$$

$$(R(\underset{\sim}{T}))_{1,4} = [(R(\underset{\sim}{C}))_{1,4} - (R(\underset{\sim}{T}))_{1,6}(R(\underset{\sim}{T}))_{2,6} - (R(\underset{\sim}{T}))_{1,11}(R(\underset{\sim}{T}))_{2,11}]/((R(\underset{\sim}{T}))_{2,4})$$

$$= [42 - 2 \times 4 - 1 \times 2]/8 = 4 \ .$$

In a similar way (A6) and (A7), used in calculating $\underset{\sim}{F}$, can be rewritten as

$$(R(\underset{\sim}{F}))_{g_j,j} = ((R(\underset{\sim}{T}))_{g_j,j})^{-1} \qquad j = 1, \ldots, n$$

and

$$(R(\underset{\sim}{F}))_{rj} = -[\sum_{k=r}^{g_j-1} (R(\underset{\sim}{F}))_{rq}(R(\underset{\sim}{T}))_{kj}](R(\underset{\sim}{F}))_{g_j,j}$$

where $q = N_{kj}$, j s.t. $g_j = s$, $r < s \leq t$

for instance

$$(R(\underset{\sim}{F}))_{1,11} = -[((R(\underset{\sim}{F}))_{1,1}(R(\underset{\sim}{T}))_{1,11} + (R(\underset{\sim}{F}))_{1,4}(R(\underset{\sim}{T}))_{2,11} + (R(\underset{\sim}{F}))_{1,6}(R(\underset{\sim}{T}))_{3,11}](R(\underset{\sim}{F}))_{4,11}$$

$$= -[(2/16) \times 1 + (1/16) \times 2 + 0 \times 4] \times (2/16) = 0 .$$

APPENDIX D - FORTRAN LISTING OF PROGRAM TO CARRY CALCULATIONS IN APPENDIX C

```
      DIMENSION IB(5,100),CB(5,100),NF(5),NT(5),NG(100)
C
C        THIS PROGRAM INVERTS THE SUB-MATRIX Z'Z+GAMMA(-1) AS DESCRIBED
C        IN THE APPENDIX OF THE PAPER
C        'THE ESTIMATION OF HERITABILITY WITH UNBALANCED DATA(2)'
C        BY R.THOMPSON.
C
C     FORMAT STATEMENTS
100   FORMAT(11I3)
101   FORMAT(11F9.4)
103   FORMAT(2I4)
C
C        READ IN THE NUMBER OF INDIVIDUALS(N)
C        AND THE NUMBER OF GENERATIONS(TT) IN I4 FORMAT.
C
      READ(5,103)NB,IG
      IGM1=IG-1
      IG1=IG+1
      IG2=IG+2
      DO 2 I=1,NB
      DO 1 L=1,IG1
      CB(L,I)=0
1     IB(L,I)=0
2     IB(IG2,I)=0
C
C        THE MATRIX IB IS OF DIMENSION (TT+2)*N.
C        THE MATRIX HAS BEEN INITIALISED AND THE FIRST TWO ROWS
C        ARE NOW READ IN.
C        THE FIRST ROW IS THE N INDIVIDUALS (J=1.....N)
C        AND THE SECOND ROW ARE THEIR FATHERS. A COLUMN REPRESENTS
C        THE JTH INDIVIDUAL(ROW1) AND HIS FATHER(ROW2).
C        IF A FATHER IS NOT KNOWN INPUT '0'.
C        INPUT THE TWO ROWS IN 11I3 FORMAT.
C
      READ(5,100)((IB(J,I),I=1,NB),J=1,2)
C
C        IN THE VECTOR NG(DIMENSION N)INPUT THE
C        DIAGONAL ELEMENTS OF THE INVERSE OF GAMMA.
C        INPUT THE VECTOR IN 11I3 FORMAT.
C
      READ(5,100)(NG(I),I=1,NB)
      DO 4 L=2,IG
      L1=L+1
      DO 5 I=1,NB
      IF(IB(L,I))6,5,6
6     I1=I-1
      DO 7 J=1,I1
      IF(IB(1,J)-IB(L,I)) 7,8,7
8     IB(L1,I)=IB(2,J)
      GOTO 5
7     CONTINUE
5     CONTINUE
4     CONTINUE
      DO 10 I=1,NB
      IQ=IG
      DO 11 L=1,IG
      IF(IB(IG1,I))10,13,10
```

```
13    IQ=IQ-1
      DO 14 L1=1,IG
      L2=IG2-L1
      L3=L2-1
14    IB(L2,I)=IB(L3,I)
      IB(1,I)=0
11    CONTINUE
10    IB(IG2,I)=IQ
      DO 26 L=1,NB
      DO 25 I=1,IG1
25    NF(I)=IB(I,L)
      I1=IG1
      DO 26 I=1,IG1
      IB(I,L)=NF(I1)
26    I1=I1-1
C
C        THE FIRST (TT+1) ROWS OF IB IS THE MATRIX N(EQUATION(A9)).
C        THE LAST ROW INDICATES THE NUMBER OF THE GENERATION OF
C        THE JTH INDIVIDUAL(GJ-1)(TOP OF PAGE A7).
C
      DO 18 I=1,NB
18    NT(IB(IG2,I)+1)=I
      DO 19 I=1,IG
19    NF(I+1)=NT(I)+1
      NF(1)=2
C
C        THE DIAGONAL ELEMENTS OF Z'Z ARE NOW CALCULATED
C        ACCORDING TO EQUATION(A14).
C
      DO 20 I=1,IG1
      JF=NF(I)
      DO 20 J=JF,NB
20    CB(I,IB(I,J))=CB(I,IB(I,J))+4**(IG+I-1-IB(IG2,J))
C
C        THE DIAGONAL ELEMENTS OF Z'Z HAVE NOW BEEN FORMED.
C
C        THE ELEMENTS OF THE INVERSE OF GAMMA ARE NOW ADDED
C        TO THE MATRIX Z'Z.
C
      DO 50 J=1,IG1
      DO 50 I=1,NB
      IF(CB(J,I).EQ.0.0)GOTO 50
      CB(J,I)=CB(J,I)+NG(I)
50    CONTINUE
C
C        THE OFF DIAGONAL ELEMENTS OF Z'Z ARE NOW CALCULATED
C        ACCORDING TO EQUATION (A15).
C
      NF(1)=1
      DO 21 I=2,NB
      K=IB(IG2,I)
      DO 21 J=1,K
21    CB(J,I)=CB(K+1,I)/(2**(K+1-J))
C
C        THE MATRIX R(C) HAS NOW BEEN FORMED. (EQUATION(A11))
C
C        THE MATRIX R(T) IS NOW CALCULATED
C        ACCORDING TO EQUATIONS ON PAGE A8.
C
```

```
      I1=IG1
      DO 30 I=1,IG1
      IF(I1-IG1) 31,32,31
31    JF=NF(I1+1)
      DO 33 J=JF,NB
      L=IB(I1,J)
      DO 34 K=1,I1
34    CB(K,L)=CB(K,L)-CB(I1,J)*CB(K,J)
33    CONTINUE
32    JF=NF(I1)
      JT=NT(I1)
      DO 35 J=JF,JT
      CB(I1,J)=SQRT(CB(I1,J))
      IF(I1-1)  36,35,36
36    K1=I1-1
      DO 37 K=1,K1
37    CB(K,J)=CB(K,J)/CB(I1,J)
35    CONTINUE
30    I1=I1-1
C
C        THE MATRIX R(T) HAS NOW BEEN FORMED. (EQUATION(A12))
C
C        THE RECIPROCAL OF THE DIAGONAL ELEMENTS ARE NOW CALCULATED
C        ACCORDING TO EQUATION (A6).
C
      I1=IG1
      DO   40 I=1,IG1
      JF=NF(I)
      JT=NT(I)
      DO 40 J=JF,JT
40    CB(I,J)=1./CB(I,J)
C
C        THE FINAL MATRIX R(F) IS NOW CALCULATED(A(13)).
C
      DO 42 I=1,IG
      IM1=I-1
      IP1=I+1
      DO 43  K=IP1,IG1
      KM1=K-1
      JF=NF(K)
      JT=NT(K)
      DO 44 J=JF,JT
44    CB(I,J)=CB(I,J)*CB(I,IB(I,J))
      IF(I-KM1) 46,45,46
46    DO 47  J=JF,JT
      DO 47 L=IP1,KM1
47    CB(I,J)=CB(I,J)+CB(L,J)*CB(I,IB(L,J))
45    DO 48 J=JF,JT
48    CB(I,J)=-CB(I,J)*CB(K,J)
43    CONTINUE
42    CONTINUE
      WRITE(6,200)
200   FORMAT(' MATRIX R(F)')
C
C*****WARNING************
C        TO OUTPUT MATRIX IN THE FOLLOWING FORM,
C        THE CARRIAGE WIDTH NEEDS TO BE AT LEAST 9*N CHARACTERS
      WRITE(6,101)((CB(J,I),I=1,NB),J=1,IG1)
      STOP
      END
```


ACKNOWLEDGEMENT

I would like to acknowledge the careful typing of the continual drafts by Mrs Norma Linn.

21

Estimation of quantitative genetic parameters

by

R. Thompson

Estimation of quantitative genetic parameters

Robin Thompson
A.R.C. UNIT OF STATISTICS
THE KING'S BUILDINGS, EDINBURGH EH9 3JZ, SCOTLAND

1. INTRODUCTION

The genetic parameters that will be considered are genetic variances and covariances. Standard references (for instance Cockerham [1963], Falconer [1960] and Kempthorne [1957]) have discussed the common methods and problems associated with estimating these parameters. Typically some system of mating is used to generate sets of relatives raised in one or more environments. Often an analysis of variance (for collateral relatives) or covariance (for non-collateral relatives) based on the mating and environmental design can be easily constructed. The resulting variance and covariance components are usually easily interpreted in terms of covariances between relatives. These covariances between relatives can be also interpreted in terms of genetic and environmental components and hence estimates of genetic variance can be derived. The key role of the analysis of variance is not

Proceedings of the International Conference on QUANTITATIVE GENETICS
© Iowa State University Press. Ames, Iowa. 1977.

surprising since it is a way of partitioning variance, and genetic variances arose out of partitioning phenotypic variance (Fisher [1918]).

In many cases (in fact most of those discussed in standard texts) this partition of variance is enough to make estimation simple and efficient. However there are cases when this is not so and in this paper we discuss maximum likelihood (ML) methods in some of these cases. There are two main cases. One case is the balanced designs where a partition of variance is possible but there are more covariances between relatives than parameters to estimate and hence for some parameters more than one estimate can be derived. In Section 2 ML estimation is discussed and in particular a simple estimation procedure very similar to weighted least squares is given.

The other case considered is the unbalanced designs, which can occur, for example, when it is impossible to raise families of equal size. In this situation many estimation procedures for variance components have been suggested (Searle [1971]). These are usually based on analogies with the analysis of variance for balanced data. In the past, ML estimation has been rarely attempted primarily because of the computational difficulties. It is argued in Section 3 that, in some cases at least, the computational difficulties are no worse than in many ad hoc schemes and that the terms in the ML estimating equations are often useful in animal breeding studies.

In Section 4 we discuss the modifications needed when animals used as parents are selected on their phenotypic performance. In Section 5 we discuss the case when data are available from more than 2 discrete generations.

2. ESTIMATION FROM BALANCED DESIGNS

The case when a balanced design generates more covariances between relatives than parameters is now discussed. For example consider a hierarchical structure in which there are s sires,

d dams mated to each sire and n offspring raised from each dam, and data available on offspring and parents (Hill and Nicholas [1974], Thompson [1976]). This design generates five variances and covariances between relatives, namely, covariances between full sibs, σ_{fs}, between half sibs, σ_{hs}, between father and offspring, σ_{fo}, between mother and offspring, σ_{mo}, and phenotypic variance σ_p^2. In heritability estimation these structural parameters are often interpreted in terms of three environmental and genetic parameters: phenotypic variance, additive genetic variance, σ_A^2, and σ_K^2, the part of the full sib covariance not due to additive variance, which therefore contains dominance and common environmental terms. The relationship between the two sets of parameters is given in Table I.

TABLE I

Covariances among Relatives in terms of σ_A^2, σ_K^2 and σ_p^2.

Covariance	σ_A^2	σ_K^2	σ_p^2
σ_{fs}	1/2	1	0
σ_{hs}	1/4	0	0
σ_{fo}	1/2	0	0
σ_{mo}	1/2	0	0
σ_p^2	0	0	1

We will use this model for illustration even though there are several assumptions about genetic relationships, for instance no epistasis and no maternal effects, that might not be appropriate. We see there are 5 covariances between relatives and 3 parameters to estimate. Hill and Nicholas [1974] have shown how the parent-offspring and half sib estimates can be combined by evaluating the variances and covariances of the estimates. This is tedious and difficult to generalize. We now know how the ML estimates can be conveniently calculated.

Suppose X_i, Y_{ij} and Z_{ijm} represent observations on sires, dams and offspring and these are normally distributed about means μ_X, μ_Y and μ_Z, with covariances between observations given in

Table I. A convenient way of summarizing the data is to calculate 3 sum of squares and products matrices representing variation within dams ($\underset{\sim}{S}_1$) between dams within sires ($\underset{\sim}{S}_2$) and between sires ($\underset{\sim}{S}_3$). We let $\bar{x}_{\cdot}$, $\bar{y}_{i\cdot}$ etc. denote means taken over the subscript replaced by a dot and

$$z_{ijm} = z_{ijm} - \bar{z}_{ij\cdot}, \quad z_{ij\cdot} = \bar{z}_{ij\cdot} - \bar{z}_{i\cdot\cdot}, \quad z_{i\cdot\cdot} = \bar{z}_{i\cdot\cdot} - \bar{z}_{\cdot\cdot\cdot},$$

$$y_{ij} = y_{ij} - \bar{y}_{i\cdot}, \quad y_{i\cdot} = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}, \quad x_i = x_i - \bar{x}_{\cdot}.$$

The sum of squares and products matrices are:

$$\underset{\sim}{S}_1 = \left(\sum_i \sum_j \sum_m z^2_{ijm}\right), \quad \underset{\sim}{S}_2 = \begin{pmatrix} n\sum_i \sum_j z^2_{ij\cdot} & \sum_i \sum_j z_{ij\cdot} y_{ij} \\ \sum_i \sum_j z_{ij\cdot} y_{ij} & \sum_i \sum_j y^2_{ij} \end{pmatrix},$$

$$\underset{\sim}{S}_3 = \begin{pmatrix} nd\sum_i z^2_{i\cdot\cdot} & d\sum_i z_{i\cdot\cdot} y_{i\cdot} & \sum_i z_{i\cdot\cdot} x_i \\ d\sum_i z_{i\cdot\cdot} y_{i\cdot} & d\sum_i y^2_{i\cdot} & \sum_i y_{i\cdot} x_i \\ \sum_i z_{i\cdot\cdot} x_i & \sum_i y_{i\cdot} x_i & \sum_i x^2_i \end{pmatrix}. \tag{1}$$

We note that the z, y and x squared terms represent terms in the analysis of variance of offspring, dam and sire measurements respectively and the cross product terms represent terms in the analysis of covariance. The degrees of freedom associated with $\underset{\sim}{S}_1$, $\underset{\sim}{S}_2$ and $\underset{\sim}{S}_3$ are $\nu_1 = sd(n-1)$, $\nu_2 = s(d-1)$ and $\nu_3 = s-1$. The expected value of $\underset{\sim}{S}_h$ denoted by $\nu_h \underset{\sim}{V}_h$ can be written in terms of the covariances between relatives. We find

$$\underset{\sim}{V}_1 = (\sigma^2_p - \sigma_{fs}), \quad \underset{\sim}{V}_2 = \begin{pmatrix} \sigma^2_p - \sigma_{fs} + n(\sigma_{fs} - \sigma_{hs}) & \sigma_{mo} \\ \sigma_{mo} & \sigma_p \end{pmatrix},$$

$$\underset{\sim}{V}_3 = \begin{pmatrix} (\sigma^2_p - \sigma_{fs} + n(\sigma_{fs} - \sigma_{hs}) + nd\sigma_{hs} & \sigma_{mo} & \sigma_{fo} \\ \sigma_{mo} & \sigma^2_p & 0 \\ \sigma_{fo} & 0 & \sigma^2_p \end{pmatrix}. \tag{2}$$

The likelihood of all the data can be partitioned into two parts, one due to the fixed effects and one due to error contrasts i.e. contrasts with expectation independent of the fixed effects.

We use this latter log-likelihood, $\mathcal{L}$, to estimate the variance parameters arguing that in the absence of knowledge about the fixed effects the former provide no information about the variance parameters (Patterson and Thompson [1971]). In this example the $\mathcal{L}$ is equivalent to the log-likelihood of $\underset{\sim}{S}_1$, $\underset{\sim}{S}_2$ and $\underset{\sim}{S}_3$ and can be written as

$$\mathcal{L} = \text{const} - \frac{1}{2} \sum_{h=1}^{3} \nu_h \left[\log|\underset{\sim}{V}_h| + \text{tr}(\underset{\sim}{M}_h \underset{\sim}{V}_h^{-1})\right] \qquad (3)$$

where $\underset{\sim}{M}_h = \underset{\sim}{S}_h/\nu_h$. In order to differentiate (3) with respect to the parameters we express the $\underset{\sim}{V}$'s as a linear function of the parameters, i.e.

$$\underset{\sim}{V}_h = \underset{\sim}{X}_{hA}\,\sigma_A^2 + \underset{\sim}{X}_{hK}\sigma_K^2 + \underset{\sim}{X}_{hp}\sigma_p^2 \qquad (4)$$

where the $\underset{\sim}{X}$'s are known matrices. The matrices $\underset{\sim}{X}_{hi}$ (h=1,2,3; i=A,K,p) can be derived by replacing the covariances in $\underset{\sim}{V}_h$ by the corresponding coefficients for σ_i^2 in Table I, for instance $\underset{\sim}{X}_{1A} = (\frac{1}{2})$, $\underset{\sim}{X}_{1K} = (1)$ and $\underset{\sim}{X}_{1p} = 1$. The values of σ_i^2 that maximize (3) satisfy

$$\frac{\partial \mathcal{L}}{\partial \sigma_i^2} = \sum_{h=1}^{3} \nu_k \,\text{tr}(\underset{\sim}{V}_h^{-1}\underset{\sim}{M}_h\underset{\sim}{V}_h^{-1}\underset{\sim}{X}_{hi}) - \sum_{h=1}^{3} \nu_K \,\text{tr}(\underset{\sim}{V}_h^{-1}\underset{\sim}{X}_{hi}) = 0 \quad . \qquad (5)$$

Usually (5) cannot be solved explicitly and an iterative solution is needed. One based on using the expected values of the second differentials that is very similar to weighted least squares is suggested by Anderson [1973]. In this scheme $\hat{\sigma}_i^2$ is estimated from

$$\sum_{j=1}^{3} \tilde{\underset{\sim}{A}}_{ij}\,\hat{\sigma}_j^2 = \tilde{\underset{\sim}{B}}_i \qquad (6)$$

where $\tilde{\underset{\sim}{A}}_{ij} = \sum_{h=1}^{3} \nu_h \,\text{tr}(\tilde{\underset{\sim}{V}}_h^{-1}\underset{\sim}{X}_{hi}\tilde{\underset{\sim}{V}}_h^{-1}\underset{\sim}{X}_{hj})$, $\tilde{\underset{\sim}{B}}_i = \sum_{h=1}^{3} \nu_h \,\text{tr}(\tilde{\underset{\sim}{V}}_h^{-1}\underset{\sim}{X}_{hi}\tilde{\underset{\sim}{V}}_h^{-1}\underset{\sim}{M}_h)$ (7)

and $\tilde{\underset{\sim}{V}}_h$ is an initial estimate of $\underset{\sim}{V}_h$. The procedure is repeated using $\hat{\sigma}_i^2$ to give $\tilde{\underset{\sim}{V}}_h$ (from (4)) until the estimates converge. The relationship with weighted least squares becomes apparent if we consider the linear model

$m_h = \sum_{i=1}^{q} x_{hi}\,\theta_i + e_h$ (h = 1, ..., H) with the e's uncorrelated,

with variances w_h. The weighted least squares estimates of θ_i satisfy

$$\sum_{j=1}^{q} A_{ij} \hat{\theta}_j = B_i \qquad (8)$$

where $A_{ij} = \sum_{h=1}^{H} w_h^{-1} x_{hi} x_{hj}$ and $B_i = \sum_{h=1}^{H} w_h^{-1} x_{hi} m_h$. (9)

Obviously the weight given in (6) to $\underset{\sim}{M}_h$ depends on ν_h and $\tilde{\underset{\sim}{V}}_h$. If the $\underset{\sim}{M}_h$ are scalars then the weights are inversely proportional to v_h^2/ν_h which is not surprising since then $\underset{\sim}{M}_h$ has a χ^2 distribution with mean $\underset{\sim}{V}_h$ and degrees of freedom ν_h. This procedure has been introduced using the hierarchical example but can be used whenever the data can be split into independent sum of squares and product matrices and their expectation is a linear function of variance parameters. Other analogies with least squares carry over; if $\underset{\sim}{A}$ is singular not all the parameters can be estimated, $2\hat{\underset{\sim}{A}}^{-1}$ gives the asymptotic variance-covariance matrix of the estimates which makes it relatively easy to compare alternative designs and the efficiency of the ML versus other estimation procedures.

Equations similar to (6) and (7) arise in other estimation procedures. For example they occur in minimum norm quadratic unbiased estimation (MINQUE) (Rao [1973]) if $\tilde{\underset{\sim}{V}}_h$ is chosen to correspond with the norm being minimized. Other methods (for instance Horn, Horn and Duncan [1975]) follow from replacing $\tilde{\underset{\sim}{V}}_h$ by $\underset{\sim}{V}_h$ in part of (7) and manipulating (6). Another possibility is to use weighted least squares on the elements of $\underset{\sim}{M}_h$ or the covariances between relatives (Hayman [1960]). This leads to the same estimates as in the ML procedure but needs the derivation of variances of and covariances between the elements of $\underset{\sim}{M}_h$ or the covariances between relatives. Although, in theory, these can be found, in practice the calculations can be intractable. Eisen [1967] suggested a design for estimating maternal genetic variances that generated 13 covariances between relatives and so $13 \times 14/2 = 91$ variances and covariances would be needed to implement the weighted least squares procedure.

3. ESTIMATION IN UNBALANCED DESIGNS

We discuss in this section ML estimation in unbalanced designs. In unbalanced designs sensible partitions of the data are not as obvious as in balanced designs. Often linear models for the observations (as opposed to linear models for the variance parameters as in Section 2) are introduced to generate appropriate partitions. A simple two factor model will be used to illustrate the main points. Extensions to more general models follow naturally but need matrix algebra to express the results compactly.

We assume a linear model of the form

$$y_{ki\ell} = \alpha_k + b_i + e_{ki\ell} \quad . \tag{10}$$

In sire evaluation this model is often used and then $y_{ki\ell}$ is the yield of the ℓ-th daughter of sire i in herd-season k, α_k is the effect of herd-season k and $e_{ki\ell}$ is a random variable normally distributed with mean zero and variance σ^2. If no other assumptions are made about α_k and b_i then α_k and b_i are called fixed effects and (10) is a fixed effects model. Alternatively if we assume the b_i are normally distributed with mean zero and variance σ_b^2 then the b_i are called random effects and (10) is then a mixed effects model. This mixed model implies that $\text{var}(y_{ki\ell}) = \sigma^2 + \sigma_b^2$ and $\text{cov}(y_{ki\ell}, y_{k'i'\ell'}) = \sigma_b^2$ if $i = i'$ and $= 0$ if $i \neq i'$. So the model can be written

$$y_{ki\ell} = \alpha_k + e'_{ki\ell} \tag{11}$$

where $e'_{ki\ell}$ has variance $\sigma^2 + \sigma_b^2$ and $\text{cov}(e'_{ki\ell}, e'_{k'i'\ell'}) = \sigma_b^2$ if $i = i'$ and 0 if $i \neq i'$. Often (11) is a covenient way of thinking about genetic models and helps in formulating the linear model (10). In the sire evaluation case if the covariances between daughters of a bull are assumed to be $\sigma_A^2/4$ and the variance of an observation is σ_p^2 this is consistent with a mixed model with $\sigma_b^2 = \sigma_A^2/4$ and $\sigma^2 = \sigma_p^2 - \sigma_A^2/4 = \sigma_p^2(1-h^2/4)$, where $h^2 = \sigma_A^2/\sigma_p^2$.

Searle [1971] has reviewed methods for estimating σ_b^2 and σ^2. Most follow the simple recipe of equating two sums of

squares to their expectations. One of the commonest (called the method of fitting constants or Henderson's method 3) is now outlined because the development is useful in understanding ML estimation. If α_k and b_i were fixed effects and were estimated by least squares they would satisfy

$$n_{kO}\,\hat{\alpha}_k + \sum_i n_{ki}\,\hat{b}_i = y_{kO} \quad , \tag{12}$$

$$\sum_k n_{ki}\,\hat{\alpha}_k + n_{Oi}\,\hat{b}_i = y_{Oi} \quad , \tag{13}$$

where n_{ki} is the number of daughters of sire i in herd-season k and O indicates summation over a suffix. An analysis of variance can be constructed:

Source	Sum of squares	
Herd-seasons	$\sum y_{kO}^2/n_{kO}$	
Sires (adjusted for Herd-seasons)	$\sum \hat{b}_i\, y_{Oi}$	(14)
Residual	$\sum y_{ki\ell}^2 - \sum \hat{\alpha}_k\, y_{kO} - \sum \hat{b}_i\, y_{Oi}$	(15)

The sires sum of squares is the difference between fitting a model with α_k and b_i and with α_k. The residual sum of squares is the sum of squares of deviations $(y_{kij} - \hat{\alpha}_k - \hat{b}_i)$. In the method of fitting constants (14) and (15) are equated to their expectation, which are functions of σ^2 and σ_b^2, and hence σ^2 and σ_b^2 can be estimated. The efficiency of this procedure is in general unknown and depends on the degree of unbalance and relative magnitude of σ^2/σ_b^2. However using this method with some unbalanced designs more precise estimates can be obtained if some of the data are removed (Swiger, Harvey, Everson and Gregory [1964]). Another misgiving I feel is in the ambivalence in the role of b_i. They are first assumed to be fixed effects to generate the sum of squares (14) and (15) and then assumed to be random effects to calculate their expectations. Further, if the mixed model is interpreted as (11) i.e. as a model for α_k with correlated errors it can be argued that weighted least squares and not least squares should be used to estimate α_k. Weighted least squares would usually require the inversion of a matrix of size the number of observations. However in mixed models of this

type Henderson (in Henderson, Kempthorne, Searle and Von Krosigk [1959]) has shown that this inversion can be eliminated and that the weighted least squares estimate of α_k satisfies

$$n_{kO}\,\tilde{\alpha}_k + \Sigma_i\, n_{ki}\,\tilde{\beta}_i = y_{kO} \quad , \tag{16}$$

$$\Sigma_k\, n_{ki}\,\tilde{\alpha}_k + (n_{Oi} + \gamma^{-1})\,\tilde{\beta}_i = y_{Oi} \quad , \tag{17}$$

where $\gamma = \sigma_b^2/\sigma^2$. These equations are very similar to (12) and (13) except that the coefficient n_{Oi} in (13) is replaced by $(n_{Oi} + \gamma^{-1})$ in (17). Henderson [1973] has emphasized that the $\tilde{\beta}_i$ can be interpreted as the predicted breeding values of bull i . In the sire evaluation case $\tilde{\beta}_i = [n_{Oi}\, h^2/(4 + (n_{Oi} - 1)h^2)]\,(y_{Oi} - \Sigma_k\, n_{ki}\,\tilde{\alpha}_k)/n_{Oi}$ i.e. the mean daughter-yield corrected for herd seasons is regressed back by a factor $[n_{Oi}\, h^2/(4 + (n_{Oi} - 1)h^2)]$.

Further $\tilde{\beta}_i$ plays a key part in the ML estimation of σ^2 and σ_b^2 . Patterson and Thompson [1971] have shown that the ML estimating equations are equivalent to equating $\Sigma y_{kij}^2 - \Sigma\tilde{\alpha}_k\, y_{kO} - \Sigma\tilde{\beta}_i\, y_{Oi}$ and $\Sigma\,\tilde{\beta}_i^2$ to their expected value. The first term is similar to a sum of squares of residuals (cf (15)), and the second term is the sum of squares of bull's predicted values. Once again an iterative scheme is usually needed to estimate σ^2 and σ_b^2 since (17) depends on σ_b^2/σ^2 . In practical cases I have found the iterative scheme outlined by Patterson and Thompson [1971] has converged in two or three iterations. The fitting constant method essentially gives equal weight to each observation, other methods give weights to the family means that are functions of their size. For the one-way classification i.e. only one herd-season $\Sigma\,\tilde{\beta}_i^2$ can be written as $\Sigma\,\gamma^2\,[\gamma + n_{Oi}^{-1}]^{-2}\,[(y_{Oi} - n_{Oi}\,\tilde{\alpha}_1)/n_{Oi}]^2$ and $[\gamma + n_{Oi}^{-1}]^{-2}$ is the weighting suggested by Robertson [1962].

The ML method can be extended to more complicated cases.

Thompson [1977] has considered the unbalanced version of the hierarchical design discussed in Section 2. Kempthorne and Tandon [1953] (for a single family classification) and Ollivier [1974] (for a hierarchical family classification) have suggested regression schemes weighting families according to size in order to make most use of the parent-offspring information. The ML method automatically does this and, where appropriate, uses the extra information from the sib covariances and enables fixed effects to be estimated.

The ML method can also be generalized to deal with q traits. In the two factor case we estimate $q \times q$ matrices $\underset{\sim}{\sigma}^2$ and $\underset{\sim}{\sigma}^2_b$. Useful equations are then (Thompson [1973])

$$n_{kO}\,\tilde{\alpha}_{km} + \sum_i n_{ki}\,\tilde{\beta}_{im} = y_{kOm} \qquad (18)$$

$$\sum_k n_{ki}\,\tilde{\alpha}_{km} + n_{Oi}\,\tilde{\beta}_{im} + \sum_{k=1}^{q} \gamma^{-1}_{mk}\,\tilde{\beta}_{ik} = y_{Oim} \qquad (19)$$

where $\underset{\sim}{\gamma}^{-1} = \underset{\sim}{\sigma}^2(\underset{\sim}{\sigma}^2_b)^{-1}$ and the suffix m represents the m-th trait ($m = 1, \ldots, q$). Again $\tilde{\beta}_{im}$ can be interpreted as the predicted value for sire i for trait m. It is equivalent to combining the data on all q traits, corrected for herd-season effects, by means of a selection index to give the predicted value for the m-th trait. The ML estimating equations are natural extensions of the univariate equations. For instance, the sum of squares and products of the values $\tilde{\beta}_{im}$ are used. If $\underset{\sim}{\gamma}^{-1}$ is diagonal, then (18) and (19) separate into q parts each like (16) and (17) and there is no connection between the q variates.

If $\underset{\sim}{\gamma}^{-1}$ is not diagonal a canonical transformation of the variates enables the equations to be solved in q parts. The q new derived variates are

$$y^*_{kijr} = \sum_m T_{rm}\,y_{kijm} \qquad (m = 1, \ldots, q)$$

where $\underset{\sim}{T}$, the matrix of coefficients T_{rm} satisfies

$$\underset{\sim}{T}\,\underset{\sim}{\sigma}^2_b\,\underset{\sim}{T}' = \underset{\sim}{I} \quad \text{and} \quad \underset{\sim}{T}\,\underset{\sim}{\sigma}^2\,\underset{\sim}{T}' = \underset{\sim}{D}$$

where $\underset{\sim}{D}$ is a diagonal matrix. These canonical variates sometimes have a genetic interpretation in terms of which linear combination

of traits is most heritable (Rouvier [1969]). They might also be useful in interpreting results on the effects of errors in parameter estimates on the efficiency of selection indices (Harris [1964]).

4. SELECTION OF PARENTS

Sometimes, either through design or accident, the animals that are used as parents are chosen on their phenotypic performance. Then some of the usual methods of estimation are biased, for example heritability if estimated by sib covariances, genetic correlations if estimated by parent-offspring regression. In this section it is argued that these difficulties are removed if ML is used.

Suppose we have observations on $\nu_1 + \nu_2$ parents (y_i) and on ν_2 offspring (z_i). Suppose also y_{1i} and y_{2i} are normally distributed with mean zero and variance V_1 and V_{22} and covariances V_{12} between y_i and z_i and also that parents are chosen at random. Let S_1 be the sum of squares for the parental data and $\underset{\sim}{S}_2$ the sum of squares and cross products matrix for parent and offspring data. Let $\nu_2\underset{\sim}{V}_2$ be the expected value of $\underset{\sim}{S}_2$, then $\underset{\sim}{V}_2$ and $\underset{\sim}{S}_2$ can be partitioned as

$$\underset{\sim}{V}_2 = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \qquad \underset{\sim}{S}_2 = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

where $V_{11} = V_1$ and $V_{21} = V_{12}$. The log-likelihood can be written as

$$\mathcal{L} = \text{const} - \tfrac{1}{2}(\nu_1 \log|V_1| + \text{tr}[(S_1 - S_{11})V_1^{-1}] + \nu_2 \log|V_2| + \text{tr}(\underset{\sim}{S}_2\underset{\sim}{V}_2^{-1})) \quad . \tag{20}$$

$\mathcal{L}$ is of the same form as (3) and hence (6) can be used if V_1 and $\underset{\sim}{V}_2$ are linear functions of the unknown parameters. An alternative instructive form for (20) follows if we partition $\mathcal{L}$ into two independent parts, one part, $\mathcal{L}_1$, the log-likelihood of y_{1i}, the parental data and another part, $\mathcal{L}_2$ the log-likelihood of $z_i - V_{21}V_{11}^{-1}y_i$ which can be thought of as the offspring record given (or conditional on) the parental record. Defining $S_{22.}$ and

$V_{22.}$ as the sum of squares and the variance of $z_i - V_{21}V_{11}^{-1}y_i$, we find $\mathcal{L}_1$ and $\mathcal{L}_2$ can be written as

$$\mathcal{L}_1 = \text{const} - \tfrac{1}{2}((\nu_1 + \nu_2)\log|V_1| + \text{tr}(S_1V_1^{-1})) , \qquad (21)$$

$$\mathcal{L}_2 = \text{const} - \tfrac{1}{2}(\nu_2\log|V_{22.}| + \text{tr}(S_{22.}V_{22.}^{-1})) . \qquad (22)$$

We see ML essentially makes use of three pieces of information. The parental data gives information on V_1, regression of z_i on y_i gives information on $V_{21}V_{11}^{-1}$ and $z_i - V_{21}V_{11}^{-1}y_i$ gives information on $V_{22.}$.

Suppose parents are chosen on their parental values, then following Kempthorne and Von Krosigk (in Henderson et al [1959]) and Curnow [1961] we can write the log-likelihood as the log-likelihood of parental values plus the log-likelihood of offspring values given parental values. This log-likelihood is again $\mathcal{L}_1 + \mathcal{L}_2 = \mathcal{L}$ (20) and the iterative scheme in Section 2 can be used. One minor modification is needed, A_{ij}, depends on the expected values of the second moments $E(y_iz_i)$ and $E(z_i^2)$. Following Curnow [1961] we express these conditional on the selected parental values. Let $\nu_1M_{11} = S_{11}$, then noting that $z_i - V_{21}V_{11}^{-1}y_i$ has variance $V_{22.}$ and is independent of y_i we find

$$E(y_iz_i) = M_{11}V_{11}^{-1}V_{12} \qquad (23)$$

$$E(z_i^2) = V_{22.} + V_{21}V_{11}^{-1}M_{11}V_{11}^{-1}V_{12} = V_{22} - V_{21}V_{11}^{-1}(V_{11} - M_{11})V_{11}^{-1}V_{12} \qquad (24)$$

$$= V_{22} - (1 - K)V_{21}V_{11}^{-1}V_{12} \quad \text{if} \quad M_{11} = KV_{11} .$$

Using (23) and (24) A_{ij} can be found to be

$$A_{ij} = \sum_{h=1}^{2} \nu_h \, \text{tr}(V_{\sim h}^{-1}X_{\sim hi}V_{\sim h}^{-1}X_{\sim hj}) - 2 \, \text{tr}\,(V_1^{-1}(S_{11} - \nu_1V_1)V_1^{-1}X_{1i}V_1^{-1}X_{1j})$$

$$+ 2 \, \text{tr}\left[\begin{pmatrix} V_1^{-1}(S_{11} - \nu_1V_1)V_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} X_{\sim 2i}V_{\sim 2}^{-1}X_{\sim 2j}\right] . \qquad (25)$$

Terms similar to (25) have been given by Curnow [1961] (for parent-offspring data) and Thompson [1973 and 1976] (for multivariate parent-offspring data and multivariate hierarchical

structures). Using (23) and (24) it can be checked that (5) gives unbiased estimating equations for θ_i. Note that in effect we are estimating the variances in the unselected population. Covariances and variances in the selected population could be evaluated using formulae similar to (23) and (24). Equations (23) and (24) and natural extensions of them have been used to investigate the effect of selection of parents on several common estimation procedures. They can be used to justify parent-offspring regression to estimate heritability (Falconer [1960]), to give measures of biases and to suggest correction factors for sib-covariance estimates of heritability and parent-offspring estimates of genetic correlation (Reeve [1953], Brown and Turner [1968]).

The formulae (20) - (25) have been written so that they hold if y_i and z_i are vectors. Obviously V_1, V_{12}, V_{22} etc. will then be interpreted as matrices of the appropriate size. Sometimes data are only available on the selected parents and the offspring. Then maximizing the conditional likelihood of the offspring given the parents, $\mathcal{L}_2$, seems an obvious suggestion. If $\mathcal{L}_2$ is written as $\mathcal{L} - \mathcal{L}_1$ it is of the form of (3). This way of writing the log-likelihood is similar to writing, in a non-orthogonal analysis of variance, the sum of squares for factor B after adjusting for factor A as the sum of squares for factors A and B minus the sum of squares for A (Searle [1971]). Henderson [1975] has discussed estimating fixed effects and predicting random effects from unbalanced designs using a similar conditional approach.

5. MORE THAN TWO GENERATIONS

In this section we give a convenient form for the covariances between relatives in different generations in terms of the additive genetic variance. We assume for simplicity that generations are discrete so that an individual in generation t is the offspring of parents in generation $(t-1)$ and that there are N_t individuals in the t-th generation.

In each generation we order the individuals by sex with males first. Suppose we start in generation 0 and if we assume the N_0 individuals are unrelated then, the coefficients of parentage of the individuals can be represented by a $N_0 \times N_0$ matrix, $\underset{\sim}{R}_{00}$, equal to $\frac{1}{2}\underset{\sim}{I}$. We now define matrices $\underset{\sim}{Z}_t$ of size $N_{t+1} \times N_t$ relating the individuals of generation (t+1) with those in generation t in order to calculate the other coefficients of parentage. The $\underset{\sim}{Z}_t$ matrices can be written as

$$\begin{pmatrix} \underset{\sim}{Z}_{tmm} & \underset{\sim}{Z}_{tmf} \\ \underset{\sim}{Z}_{tfm} & \underset{\sim}{Z}_{tff} \end{pmatrix} \tag{26}$$

where the elements of $\underset{\sim}{Z}_t$ are either $\frac{1}{2}$ or 0. The (j, k) element of $\underset{\sim}{Z}_{tmm}$ is $\frac{1}{2}$ only if the k-th male of generation t is the father of the j-th male in generation (t + 1). The other blocks of $\underset{\sim}{Z}$ are defined similarly. Hill [1974] used similar matrices and notes that the blocks of $\underset{\sim}{Z}$ represent the alternative pathways of genes

$$\left(\begin{array}{c|c} \text{males from males} & \text{males from females} \\ \hline \text{females from males} & \text{females from females} \end{array}\right).$$

The relationship matrix for the first three generations (which indicates the general form) can now be written, if no individual is inbred, as (Thompson [1977])

$$\underset{\sim}{R} = \frac{1}{2}\begin{pmatrix} \underset{\sim}{I}_0 & \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{Z}_0 & \underset{\sim}{I}_1 & \underset{\sim}{0} \\ \underset{\sim}{Z}_1\underset{\sim}{Z}_0 & \underset{\sim}{Z}_1 & \underset{\sim}{I}_2 \end{pmatrix}\begin{pmatrix} \underset{\sim}{I}_0 & \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & \frac{1}{2}\underset{\sim}{I}_1 & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} & \frac{1}{2}\underset{\sim}{I}_2 \end{pmatrix}\begin{pmatrix} \underset{\sim}{I}_0 & \underset{\sim}{Z}'_0 & \underset{\sim}{Z}'_0\underset{\sim}{Z}'_1 \\ \underset{\sim}{0} & \underset{\sim}{I}_1 & \underset{\sim}{Z}'_1 \\ \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{I}_2 \end{pmatrix} \tag{27}$$

This is the product of a lower triangular matrix with a diagonal matrix and with an upper triangular matrix. The inverse of $\underset{\sim}{R}$, sometimes used for predicting random effects (Henderson [1976]) has a simple form since the left inverse of the lower triangular matrix is

$$\begin{pmatrix} \underset{\sim}{I}_0 & \underset{\sim}{0} & \underset{\sim}{0} \\ -\underset{\sim}{Z}_0 & \underset{\sim}{I}_1 & \underset{\sim}{0} \\ \underset{\sim}{0} & -\underset{\sim}{Z}_1 & \underset{\sim}{I}_2 \end{pmatrix}.$$

The variance matrix of the observations assuming just an additive genetic component and an environmental component is $\underset{\sim}{V} = 2\underset{\sim}{R}\,\sigma_A^2 + \underset{\sim}{I}\,\sigma_e^2$. Suppose the records in the t-th generation are $\underset{\sim}{y}_t$ and, for simplicity, these are normally distributed about a mean of 0. ML estimation of σ_A^2 and σ_e^2 depends on calculating $\underset{\sim}{V}^{-1}$ and, except for some special cases (for example observations only available on one sex (Thompson [1977])) in most practical cases this is not feasible. One suggestion is to work with deviations from parental values $\underset{\sim}{y}^{+}_{t+1} = \underset{\sim}{y}_{t+1} - \underset{\sim}{Z}_t\,\underset{\sim}{y}_t$, since the variance matrix of these deviations is tridiagonal and the covariance between deviations two or more generations apart are zero. When each dam is mated to only one sire the variance structure for $\underset{\sim}{y}^{+}_t$ corresponds to a hierarchical analysis of variance with sire and dam components σ_e^2 and covariance between full sibs $\sigma_e^2 + \sigma_A^2/2$.

Another possibility is to work with $\underset{\sim}{y}^{*}_t$, the t-th generation values conditional on, or given, the ancestors' records, which can be interpreted as deviations of actual from predicted values. For example, $\underset{\sim}{y}^{*}_1 = \underset{\sim}{y}_1 - (\sigma_A^2/(\sigma_e^2+\sigma_A^2))\,\underset{\sim}{Z}_O\underset{\sim}{y}_O$, where $(\sigma_A^2/(\sigma_e^2+\sigma_A^2))\underset{\sim}{Z}_O\underset{\sim}{y}_O$ represents a vector of mid-parent values regressed back by a factor $(\sigma_A^2/(\sigma_e^2+\sigma_A^2))$ and hence are predicted values of $\underset{\sim}{y}_1$. The terms for the next generations are more complicated, but if we approximate the variance of $\underset{\sim}{y}^{*}_1$ by $v_1\,\underset{\sim}{I}$, where v_1 is the variance of the elements of $\underset{\sim}{y}^{*}_1$, we can approximate $\underset{\sim}{y}^{*}_2$ by $\underset{\sim}{y}_2 - \underset{\sim}{Z}_1\underset{\sim}{y}_1 + (\sigma_e^2/v_1)\underset{\sim}{Z}_1\underset{\sim}{y}^{*}_1$. Using the same type of approximation and a recursive argument similar to that of Bulmer [1971] in succeeding generations we find that $\underset{\sim}{y}^{*}_{t+1}$ might be approximated by

$$\underset{\sim}{y}^{*}_{t+1} = \underset{\sim}{y}_{t+1} - \underset{\sim}{Z}_t\underset{\sim}{y}_t + (\sigma_e^2/v_t)\underset{\sim}{Z}_t\underset{\sim}{y}^{*}_t \qquad (28)$$

where $v_t = \sigma_p^2 + d_t$, $d_{t+1} = \tfrac{1}{2}d_t - \tfrac{1}{2}H_t v_t^{-1} H_t$, $d_O = 0$, $\sigma_p^2 = \sigma_A^2 + \sigma_e^2$ and $H_i = \sigma_A^2 + d_i$. As in Bulmer's case the d_i quickly converge to a limiting value d^*

$$d^* = (-2\sigma_A^2 - \sigma_p^2 + [\sigma_p^4 + 4\sigma_A^2\ (\sigma_p^2 - \sigma_A^2)]^{\frac{1}{2}})/4 \ .$$

The values of d^*/σ_p^2 for various values of σ_A^2/σ_p^2 are given in Table II and we see approximately $d^*/\sigma_p^2 = -\frac{1}{2}(\sigma_A^2/\sigma_p^2)^2$. $\sigma_p^2 + d^*$ can be thought of as the variance between the actual value and the predicted value using all parental, grand parental etc. information.

Table II

Values of d^*/σ_p^2 for various values of σ_A^2/σ_p^2 .

σ_A^2/σ_p^2	0.0	0.2	0.4	0.6	0.8	1.0
d^*/σ_p^2	-0.000	-0.030	-0.100	-0.200	-0.330	-0.500

Sometimes the regression of response on selection differential is used to estimate heritability (Falconer [1960], Hill [1971, 1972]). This is similar to putting $d_t = 0$ in (28) and working with the mean values of y_t^* rather than the individual values. The variance-covariance matrix of the mean values can be derived by arguments similar to the development of (28) which I find more appealing than Hill's [1971, 1972] intuitive genetic approach.

BIBLIOGRAPHY

ANDERSON, T.W. [1973]. Asymptotically efficient estimation of covariance matrices with linear structure. Ann. Statist. 1, 135-141.

BROWN, G.H. and TURNER, H.N. [1968]. Response to selection in Australian Merino sheep. II Estimates of phenotypic and genetic parameters for some production traits on Merino ewes and an analysis of the possible effects of selection on them. Aust. J. of Agric. Research 19, 303-22 (Corrigendum, 21, 182).

BULMER, M.G. [1971]. The effect of selection on genetic variability. Amer. Natur. 105, 201-211.

COCKERHAM, C.C. [1963]. Estimation of genetic variances. Statistical Genetics and Plant Breeding. Nat. Acad. Sci. Nat. Res. Council Publ. 982, 53-94.

CURNOW, R.N. [1961]. The estimation of repeatability and heritability from records subject to culling. Biometrics 17, 553-66.

EISEN, S.J. [1967]. Mating designs for estimating direct and maternal variances and direct-maternal covariances. Can. J. Genet. Cytol. 9, 13-22.

FALCONER, D.S. [1960]. Introduction to Quantitative Genetics. Ronald Press Co., New York, N.Y.

FISHER, R.A. [1918]. The correlation between relatives on the supposition of Mendelian inheritance. Trans. Royal Soc., Edinburgh 52, 399-433.

HARRIS, D.L. [1964]. Expected and predicted progress from index selection involving estimates of population parameters. Biometrics 20, 46-72.

HAYMAN, B.I. [1960]. Maximum likelihood estimation of genetic components of variation. Biometrics 16, 369-381.

HENDERSON, C.R. [1973]. Sire evaluation and genetic trends. In Proc. Anim. Breed. Genet. Symp. Blacksburg, Virginia pp 10-41. American Society of Animal Sciences, Champaign, Illinois.

HENDERSON, C.R. [1975]. Best linear unbiased estimation and prediction under a selection model. Biometrics 31, 423-447.

HENDERSON, C.R. [1976]. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32, 69-83.

HENDERSON, C.R., O. KEMPTHORNE, S.R. SEARLE and C.M. von KROSIGK. [1959]. The estimation of environmental and genetic trends from records subject to culling. Biometrics 15, 192-218.

HILL, W.G. [1971]. Design and efficiency of selection experiments for estimating genetic parameters. Biometrics 27, 293-311.

HILL, W.G. [1972]. Estimation of realized heritabilities from selection experiments. I Divergent selection. II Selection in one generation. Biometrics 29, 747-780.

HILL, W.G. [1974]. Prediction and evaluation of response to selection with overlapping generations. Anim. Prod. 18, 117-139.

HILL, W.G. and F.W. NICHOLAS. [1974]. Estimation of heritability by both regression of offspring on parent and intra-class correlation of sibs in one experiment. Biometrics 30, 447-468.

HORN, S.D., R.A. HORN and D.B. DUNCAN [1975]. Estimating variances in linear models. Journal of the American Statistical Association, 70, 380-385.

KEMPTHORNE, O. [1957]. An introduction to Genetic Statistics. John Wiley and Sons, New York, N.Y.

KEMPTHORNE, O. and O.B. TANDON. [1953]. The estimation of heritability by regression of offspring on parent. Biometrics 9, 90-100.

OLLIVIER, L. [1974]. La regression parent-descendant dans le cas de descendances subdivisees en familles de taille inegale. Biometrics 30, 59-66.

PATTERSON, H.D. and R. THOMPSON. [1971]. Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545-554.

RAO, C.R. [1973]. Linear Statistical Inference and its Applications (Second Edition) Wiley, New York, N.Y.

REEVE, E.C.R. [1953]. Studies in quantitative inheritance. III Heritability and genetic correlation in progeny tests using different mating systems. J. Genet. 51, 520-542.

ROBERTSON, A. [1962]. Weighting in estimation of variance components. Biometrics 18, 413-415.

ROUVIER, R. [1969]. Ponderation des valeurs genotypiques dans la selection par indice sur plusieurs caracteres. Biometrics 25, 295-308.

SEARLE, S.R. [1971]. Linear Models. Wiley, New York, N.Y.

SWIGER, L.A., W.R. HARVEY, D.O. EVERSON and K.E. GREGORY [1969]. The variance of intraclass correlation involving groups with one observation. Biometrics, 20, 818-826.

THOMPSON, R. [1973]. The estimation of variance and covariance components with an application when records are subject to culling. Biometrics 29, 527-550.

THOMPSON, R. [1976]. Design of experiments to estimate heritability when observations are available on parents and offspring. Biometrics 32, 283-304.

THOMPSON, R. [1977]. The estimation of heritability with unbalanced data. To appear in Biometrics.

22

Probabilities of non-positive definite between group or genetic covariance matrices

by

W.G. Hill and R. Thompson

BIOMETRICS 34, 429–439
September, 1978

Probabilities of Non-Positive Definite Between-Group or Genetic Covariance Matrices

W. G. HILL

Institute of Animal Genetics, West Mains Road, Edinburgh EH9 3JN, Scotland

R. THOMPSON

ARC Unit of Statistics, Mayfield Road, Edinburgh EH9 3JZ, Scotland

Summary

The probability (Q) that the estimated between-group covariance matrix is not positive definite is computed for the balanced single classification multivariate analysis of variance with random effects. It is shown that Q depends only on the roots of the matrix product of the inverse of the true within-group and the true between-group covariance matrices which, for independent variables, reduces to expressions in intra-class correlations.

Values of Q are computed for ranges of size of experiment, intra-class correlation and number of variables. Even for large experiments, Q can approach 100% if there are many variables, for example with 160 groups of size 10 and either 8 independent variables each with intra-class 0.025 or 14 variables each with intra-class correlation 0.0625. Some rationalization of the results is given in terms of the bias in the roots of the sample between-group covariance matrix.

In genetic applications, the between-group covariance matrix is proportional to the genetic covariance matrix; if non-positive definite, heritabilities and ordinary or partial genetic correlations are outside their valid limits, and the effect on selection index construction is discussed.

1. Introduction

In a one-way classification multivariate analysis of variance with random group effects, the between-group covariance component matrix is estimated from the difference between the between-group and within-group mean square and product matrices. In genetic applications this covariance matrix is used to estimate heritabilities, genetic variances, covariances and correlations and in the construction of selection indices. In these and many other applications the true between-group covariance matrix must be positive definite (p.d.), i.e. have only positive roots so that all possible linear combinations of the variables have positive variance. In some circumstances we may allow the matrix to be positive semidefinite (p.s.d.) i.e., have only zero and positive roots so that some variables are completely correlated. In the analysis of a sample of data, however, while both the matrices of between-group and within-group sums of products or mean products must be p.d. or p.s.d., their difference need not be so. It is thus possible to obtain by chance a between-group covariance matrix which has one or more negative roots. For a single variate, a negative root implies a negative variance component, intra-class correlation and, in genetic applications, heritability estimate. For two or more variates, if the between-group covariance matrix has positive diagonal elements a matrix with one or more negative roots implies ordinary or partial between-group, or genetic,

Key words: Multivariate analysis of variance; Sample covariance matrices; Genetic covariance.

correlations outside the range -1 to $+1$. Although ordinary correlations outside this range are displayed by a correlation matrix, impossible partial correlations are more easily missed.

For single variables Gill and Jensen (1968) computed the probability of obtaining negative heritability (intra-class correlation) estimates for a range of group sizes and numbers of individuals in each group, but there appears to have been no similar investigation in the multivariate case. Seal (1966, p. 177) suggests, however, that the occurrence of such non-positive definite between-group covariance matrices is the rule rather than the exception and argues against using sophisticated multivariate methods on component matrices. Seal cites examples, and, in a genetic context, Smith, King and Gilbert (1962) analysed 35 traits in a hierarchical classification with 200 half-sib families. In a detailed study of 24 of these traits they found six negative roots. In this paper we extend the results of Gill and Jensen to many variables, and particularly study the effect of increasing the number of variables. We shall specifically refer to the probability (Q) that the matrix is non-positive definite (n.p.d.), but since we deal with continuous variates the probability of obtaining a root exactly equal to zero is infinitesimally small, and we can ignore the probability that the matrix is p.s.d. Most of the analysis is for the balanced one-way classification where the estimates of components from multivariate analysis of variance (MANOVA), maximum likelihood or other methods are the same, except where restraints are put on the parameters.

2. *Model*

There is assumed to be a one-way classification with s groups each of size n, and p characters are recorded on each individual. Observations are assumed to be multivariate normally distributed with the between- and within-group effects independent of each other. The MANOVA table is as follows, with each of the matrices of dimension $p \times p$, $\boldsymbol{\Sigma}$ being p.d. and $\boldsymbol{\Psi}$ p.s.d.

Source	*df*	*SS*	*MS*	*E(MS)*
Between groups	$s-1$	$\mathbf{S}_B$	$\mathbf{B}$	$\boldsymbol{\Sigma} + n\boldsymbol{\Psi}$
Within groups	$s(n-1)$	$\mathbf{S}_W$	$\mathbf{W}$	$\boldsymbol{\Sigma}$

The matrices of sums of squares and products have independent central Wishart distributions, $\mathbf{S}_B \sim W_p[(s-1), \boldsymbol{\Sigma} + n\boldsymbol{\Psi}]$ and $\mathbf{S}_W \sim W_p\,[s(n-1), \boldsymbol{\Sigma}]$. An estimate of $\boldsymbol{\Sigma}$ is W, and of $\boldsymbol{\Psi}$ is $\hat{\boldsymbol{\Psi}} = (B - W)/n$.

There exist vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$ such that

$$\mathbf{W} = \sum_{i=1}^{p} \mathbf{v}_i\mathbf{v}_i' \text{ and } \mathbf{B} = \sum_{i=1}^{p} \nu_i\mathbf{v}_i\mathbf{v}_i' ,$$

where the roots, ν_i, satisfy the determinantal equation $|\mathbf{B} - \nu_i\mathbf{W}| = 0$ (Anderson 1958, p. 341). Thus $\hat{\boldsymbol{\Psi}}$ can be expressed as $\boldsymbol{\Sigma}_i\,(\nu_i - 1)/n\;\mathbf{v}_i\mathbf{v}_i'$, and is p.d. if all $\nu_i > 1$, $i = 1, \ldots, p$. Hence the probability that $\hat{\boldsymbol{\Psi}}$ is p.d. depends on the distribution of the smallest of the sample roots, ν_i, and so depends on the roots, λ_i, of the determinantal equation in the parameters,

$$|\boldsymbol{\Sigma} + n\boldsymbol{\Psi} - \lambda\boldsymbol{\Sigma}| = 0 \qquad (1)$$

(Anderson 1958, p. 329), which may also be written $|n\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi} - (\lambda - 1)\,\mathbf{I}| = 0$. Therefore the probability that the estimate, $\hat{\boldsymbol{\Psi}}$, of the between-group covariance matrix is p.d. depends only on the roots of $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}$ and not on the individual elements of $\boldsymbol{\Sigma}$ or $\boldsymbol{\Psi}$. Let $\boldsymbol{\lambda}$ be the vector of roots of (1).

For illustration we can therefore assume that $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ are diagonal matrices, i.e., there is no between- or within-group correlation among any of the variables, and further that $\boldsymbol{\Sigma} + \boldsymbol{\Psi}$

$= \mathbf{I}$, but we show how the results can be generalised in the discussion. Defining t_i as the intra-class correlation of variable i, $\Psi_{ii} = t_i$, $\sum_{ii} = 1 - t_i$ and from (1), $\lambda_i = 1 + nt_i/(1 - t_i)$. In genetic applications, we assume there are half-sib families and the *ith* variable has heritability h_i^2, so $t_i = h_i^2/4$ and $\lambda_i = 1 + nh_i^2/(4 - h_i^2)$. The values used for illustration are the same as those used by Gill and Jensen (1968), namely $h^2 = 0.1$, 0.25 and 0.5, corresponding to $t = 0.025$, 0.0625 and 0.125.

3. *Methods*

Since $\mathbf{B}$ and $\mathbf{W}$ are proportional to central Wishart variables, the problem of finding the probability that $\hat{\boldsymbol{\Psi}}$ is p.d. is related to that of finding the power function of the test of $\boldsymbol{\Psi} = 0$ using the smallest root of $\mathbf{W}^{-1}\mathbf{B}$. Pillai and Al-Ani (1970) have discussed this problem and developed formulae for two and three variates, applicable for all values of $\boldsymbol{\lambda}$. Their formulae are very complicated, however, so we use two other methods.

Exact method for equal roots. Venables (1974) gives an algorithm for computing $P(k_1 < x)$, where k_1 is the largest root of $\mathbf{S}_1(\mathbf{S}_1 + \mathbf{S}_2)^{-1}$ and $\mathbf{S}_1$ and $\mathbf{S}_2$ are independently Wishart distributed, $\mathbf{S}_1 \sim W_p(d_1, \boldsymbol{\theta})$, $\mathbf{S}_2 \sim W_p(d_2, \boldsymbol{\theta})$. This can be used for the case of equal roots of $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}$.

Let $\mathbf{S}_1 = s(n-1)\mathbf{W}$, so $\mathbf{S}_1 \sim W_p(s(n-1), \boldsymbol{\Sigma})$ and let $\mathbf{S}_2 = \lambda^{-1}(s-1)\mathbf{B}$ so $\mathbf{S}_2 \sim W_p(s-1, \boldsymbol{\Sigma})$, where each of the roots of (1) equals λ. Suppose k is a root of $\mathbf{S}_1(\mathbf{S}_1 + \mathbf{S}_2)^{-1}$, then $|\mathbf{S}_1 - k(\mathbf{S}_1 + \mathbf{S}_2)| = 0$ and $|\mathbf{B} - [(1-k)\,s(n-1)\lambda/(k(s-1))]\,\mathbf{W}| = 0$. Hence $(1-k)\,s(n-1)\lambda/(k(s-1))$ is the smallest root of $\mathbf{B}\mathbf{W}^{-1}$. If $\mathbf{B} - \mathbf{W}$ is n.p.d., the smallest root of $\mathbf{B}\mathbf{W}^{-1} < 1$, or, equivalently, $k_1 > [1 + (s-1)/(s(n-1)\lambda)]^{-1}$. So the probability, $Q(\boldsymbol{\lambda}, p, s, n)$, that $\mathbf{B} - \mathbf{W}$ is n.p.d. is, for equal population roots,

$$Q(\boldsymbol{\lambda}1, p, s, n) = 1 - \text{Prob}\left(k_1 < \left(1 + \frac{s-1}{s(n-1)\lambda}\right)^{-1}\right).$$

In terms of the intra-class correlation, $\lambda = 1 + nt/(1-t)$ and for $\mathbf{B} - \mathbf{W}$ p.d.,

$$Q(\lambda 1, p, s, n) = 1 - \text{Prob}\left(k_1 < \frac{s(n-1)[1 + (n-1)t]}{s(n-1)[1 + (n-1)t] + (s-1)(1-t)}\right).$$

This equation can be evaluated using Venables' algorithm only if $s - p$ is even, and requires too much computing time as s and p increase. It was used where possible, however, and where results were required for $s - p$ odd, Q was evaluated for $s-3$, $s-1$, $s+1$ and $s+3$ and the value for s obtained by non-linear interpolation. This method was checked for $p = 1$ against values obtained using the F distribution (Gill and Jensen 1968). Any discrepancies could be attributed to rounding errors in their results.

Simulation. When Venables' algorithm could not be used, Monte Carlo simulation was adopted. For each replicate experiment, the between- and within-group matrices of sums of squares and products were sampled independently from a Wishart distribution using the algorithm of Smith and Hocking (1972). $\hat{\boldsymbol{\Psi}}$ was then obtained and its roots computed, or to save machine time, a check on whether $\hat{\boldsymbol{\Psi}}$ was p.d. was made by evaluating the determinants on the main diagonal. If $\hat{\boldsymbol{\Psi}}$ is p.d. then

$$|\hat{\boldsymbol{\Psi}}_{11}| > 0, \qquad \begin{vmatrix} \hat{\boldsymbol{\Psi}}_{11} & \hat{\boldsymbol{\Psi}}_{12} \\ \hat{\boldsymbol{\Psi}}_{21} & \hat{\boldsymbol{\Psi}}_{22} \end{vmatrix} > 0, \ldots, |\hat{\boldsymbol{\Psi}}| > 0,$$

(Graybill 1961, p. 4). Roots or determinants of other submatrices were also used to find $Q(\lambda_1, \ldots, \lambda_{p'}, p', s, n)$ for $p' < p$, in addition to $Q(\boldsymbol{\lambda}, p, s, n)$ when simulation was carried out with p traits. Thus a single simulation run could provide information for all values of p', $1 \le p' \le p$.

The algorithm used to produce the Wishart distribution generated pseudorandom uniform variates, transformed these to normal deviates and then the normal deviates to chi-square using the Wilson-Hilferty approximation. As a check on the precision of the method some results for the equal root case, $Q(\lambda 1, p, s, n)$ were obtained both by simulation and Venables' algorithm. The agreement was excellent even with as few as 20 groups, and the approximation should improve as the number of groups are increased.

4. Results

Probabilities (Q) of n.p.d. matrices are given in Table 1 for a range of equal intra-class correlations (roots), number of groups and group sizes. The increase in Q as the number of traits increases can be quite startling. For example, with $s = 40$, $n = 10$ and $t = 0.0625$ ($\lambda = 5/3$), there is a less than 3% chance that a specific intra-class correlation or heritability estimate will be negative, yet a 92% chance that with 6 variables the between-group covariance matrix will not be p.s.d. The results are illustrated more strikingly in Figure 1 where a number of examples are given for 160 groups and many traits using 100 replicates. The curve of Q against number of traits is of a sigmoid form, and shows that the range of number of traits over which Q changes from essentially 0% to 100% is rather narrow.

The case of unequal intra-class correlations or roots is illustrated in Table 2. From the results of Anderson and Das Gupta (1964) it follows that Q increases monotonically with reduction in intra-class correlation. For example, with $s = 40$, $n = 5$, $p = 2$ and $t = 0.025$, 0.025 then $Q = 72\%$ (Table 1), whereas with $t = 0.025$, 0.0625, $Q = 59\%$ (Table 2) and with $t = 0.0625$, 0.0625, $Q = 43\%$ (Table 1).

5. Discussion

Rationalisation of results. At first sight the high probabilities of n.p.d. matrices are somewhat surprising and in the Appendix some qualitative justification is given for them. We show that while the sum of the roots of $\hat{\boldsymbol{\Psi}} = (\mathbf{B} - \mathbf{W})/n$ is an unbiased estimator of the sum of roots of $\boldsymbol{\Psi}$, individual roots are biased, the larger ones upwards, the smaller downward with the magnitude of the bias increasing as the number of variables increases or as the size of the experiment decreases. Using the predicted biases in the roots, approximate formulae for probabilities that $\hat{\boldsymbol{\Psi}}$ is n.p.d. are given.

Correlated variables. The results have been illustrated for independent variables, yet for any specific vector of roots, $\boldsymbol{\lambda}$, an infinite set of matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ can be represented. Examples for two variates are given in Table 3 of possible combinations of intra-class correlations, say T_1 and T_2, and corresponding between-group product moment correlations (elements of $\boldsymbol{\Psi}$ or genetic correlations, r_G) and overall product moment correlations (elements of $\boldsymbol{\Psi} + \boldsymbol{\Sigma}$ or phenotypic correlations, r_P), which have the same roots as two independent variates with $t_1 = 0.025$ and $t_2 = 0.125$. The same combinations of r_P and r_G also apply if the t_i and T_i are multiplied by a scalar, e.g. for $t_1 = 0.1$ and $t_2 = 0.5$, possible values are $T_1 = 0.2$, $T_2 = 0.4$, $(r_P, r_G) = (87\%, 92\%), (-87\%, -92\%), (0\%, 61\%), (0\%, -61\%)$. Table 3 illustrates that one of the roots may be small and thus the probability of an n.p.d. matrix high, even if neither variate has a small intra-class correlation (heritability) but they are highly correlated. With equal intra-class correlations, $t_1 = t_2$, however, it can be shown that if $r_P = r_G$, the population roots do not depend on the magnitude of r_P and r_G.

Extensions. Although the results have been given solely for the one-way classification

TABLE 1

Probability (Q%) of Obtaining a Non-Positive Definite Between-Group Covariance Matrix. Equal Roots $\lambda = 1 + nt/(1-t)$†.

n	t	p	1	2	4	6	1	2	4	6
				s = 20				s = 40		
5	0.025		40.0	78.8	99.6	100.0	33.8	71.8	98.8	100.0
	0.0625		24.4	60.7	97.6	100.0	14.7	43.1	91.2	99.8
	0.125		9.2	31.8	85.2	99.5	2.5	11.1	52.7	90.7
10	0.025		29.0	67.3	98.7	99.9	19.4	52.2	95.1	99.9
	0.0625		9.7	33.3	86.8	99.6	2.7	11.8	55.0	92.0
	0.125		1.3	7.1	44.5	88.4	0.1	0.4	5.8	29.1
20	0.025		14.3	43.8	93.0	99.9	5.4	21.0	72.3	97.5
	0.0625		1.6	8.2	48.4	90.6	0.1	0.6	7.3	43.9
	0.125		0.1	0.5	7.4	38.2	0.0	0.0	0.1	0.6
40	0.025		3.4	15.3	65.8	96.8	0.4	2.1	19.4	60.5
	0.0625		0.1	0.6	9.1	43.3	0.0	0.0	0.1	1.0
	0.125		0.0	0.0	0.3	4.4	0.0	0.0	0.0	0.0
				s = 80				s = 160		
5	0.025		26.3	62.1	97.0	100.0	17.7	48.1	91.2	99.8
	0.0625		6.3	22.7	73.0	97.0	1.4	6.5	37.3	78.2
	0.125		0.2	1.4	14.0	47.0	0.0	0.0	0.5	3.4
10	0.025		10.1	32.8	83.0	99.0	3.3	13.5	52.6	90.7
	0.0625		0.3	1.5	14.0	47.0	0.0	0.0	0.2	4.0
	0.125		0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
20	0.025		1.0	5.0	31.3	73.4	0.0	0.0	3.0	18.0
	0.0625		0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
	0.125		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
40	0.025		0.0	0.0	0.6	5.6	0.0	0.0	0.0	0.0
	0.0625		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.125		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

† Computed by Venables' algorithm, except for s = 80, 160 with p = 4,6 when simulation with 1000 replicates was used.

MANOVA, they can readily be extended to other models. In genetic applications the following hierarchical design is frequently used:

Source		*df*	*E(MS)*	
		General	*Single trait* ($\times\sigma^2$)	
Sire families	$s-1$	$\boldsymbol{\Sigma} + n\boldsymbol{\Psi} + nd\boldsymbol{\Phi}$	$1 - t_d - t_s + nt_d + ndt_s$	
Dam families/sire	$s(d-1)$	$\boldsymbol{\Sigma} + n\boldsymbol{\Psi}$	$1 - t_d - t_s + nt_d$	
Progeny/dam	$sd(n-1)$	$\boldsymbol{\Sigma}$	$1 - t_d - t_s$	

Genetic variances and covariances are usually taken from the sire components, $4\hat{\boldsymbol{\Phi}}$, estimated as the difference between the appropriate mean squares, since they are not confounded with dominance or maternal environmental effects. The probability that $\hat{\boldsymbol{\Phi}}$ is n.p.d. now depends on the roots of the determinantal equation

$$|\boldsymbol{\Sigma} + n\boldsymbol{\Psi} + nd\boldsymbol{\Phi} - \lambda(\boldsymbol{\Sigma} + n\boldsymbol{\Psi})| = 0$$

TABLE 2

Probability (Q%) of Obtaining a Non-Positive Definite Between-Group Covariance Matrix. Unequal Roots, $\lambda_1 = 1 + nt_1/(1 - t_1)$[†].

s	n	p	2			3		4		6
	t%		2.5 6.25	2.5 12.5	6.25 12.5	2.5 6.25 12.5	2.5 × 2 6.25 × 2	2.5 × 2 12.5 × 2	6.25 × 2 12.5 × 2	2.5 × 2 6.25 × 2 12.5 × 2
40	5		58.9	49.2	28.4	74.2	97.0	93.3	78.5	99.4
	10		33.4	26.3	5.0	44.0	82.5	70.6	31.1	94.4
	20		9.4	8.0	0.1	13.9	44.1	35.4	1.6	59.0
	40		0.5	0.5	0.0	1.0	5.1	4.8	0.0	9.9
160	5		28.7	21.8	2.6	34.6	77.0	63.3	14.7	85.1
	10		4.7	4.4	0.0	6.3	24.3	19.8	0.3	32.1
	20		0.0	0.0	0.0	0.1	0.4	0.6	0.0	0.7

[†] Simulated with 1000 replicates, giving sample size of 2000 for p = 2 and p = 3, and 1000 for p = 4 and p = 6.

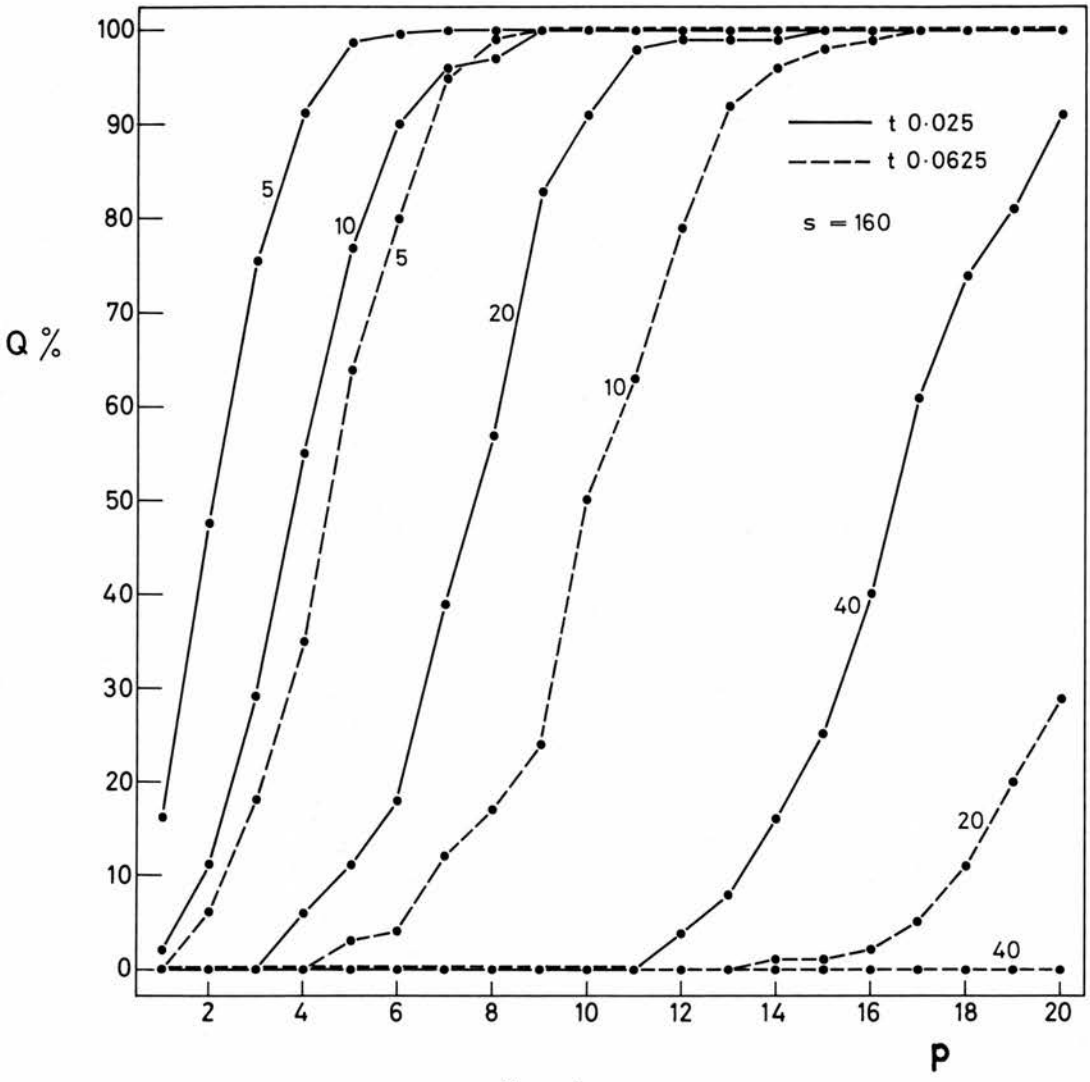


Figure 1.
Probability (Q%) of Obtaining a Non-Positive Definite Between-Group Covariance Matrix with p Traits Having Equal Intra-Class Correlation (t), for s = 160 Groups and Various Group Sizes (n).

and on the appropriate degrees of freedom (d.f.), $s - 1$ and $s(d - 1)$. For uncorrelated variables, the roots are $\lambda = 1 + ndt_s/[1 + (n - 1)t_d - t_s]$. For example, with $s = 40$, $d = 5$, $n = 8$ and $t_s = t_d = 0.025$ (corresponding to heritabilities of 10%), $\lambda = 1 + 0.5/1.15 = 1.435$. Thus the probability that $\hat{\Phi}$ is n.p.d. would be the same as in a one-way classification with $s = 40$ groups of size $n = 8$ with $\lambda = 1 + nt/(1 - t) = 1.435$, corresponding to $t = 1/19.4 = 0.0515$. For a specified total sire family size (i.e., nd) and intra-class correlation, the probability that the estimated genetic covariance matrix will be n.p.d. will be greater in the hierarchical model, partly because the d.f. in the denominator are smaller but primarily because the value of the roots, λ, are smaller, especially with large full-sib family sizes. Thus, for $n = 40$ and $t = 0.025$ in a one-way classification, $\lambda = 1 + nt/(1 - t) = 2.026$.

Gill and Jensen (1968) consider the single variable case of the hierarchical classification in

TABLE 3

Alternative Values of Between-Group (genetic, r_G) and Total (phenotypic, r_P) Correlations (%) and Intra-Class Correlations (T) Which Give the Same Roots as $t_1 = 0.025, t_2 = 0.125, r_P = 0, r_G = 0$.

T	0.025		0.050		0.075		0.100		0.125	
	$\pm r_P$	$\pm r_G$	$\pm r_P$	$\pm r_G$	$\pm r_P$	$\pm r_G$	$\pm r_P$	$\pm r_G$	$\pm r_P$	$\pm r_G$
0.025	—	—	87	61	71	41	50	25	0	0
0.050	87	61	50	−25	26	−47	0	61	50	79
0.075	71	41	97	97	0	67	26	78	71	91
0.100	50	25	87	92	97	99	50	87	87	97
0.125	0	0	50	79	71	91	87	97	—	—

more detail. They also discuss unbalanced classifications, and there is no reason to doubt that the increased probabilities of negative intra-class correlation estimates found with unbalanced classifications will carry over into the multivariate case.

Conclusions. We have shown that if an analysis with many variates is undertaken, there is a high probability that an n.p.d. between-group or genetic variance-covariance matrix will be obtained. Without specifying how the results are to be used it is not necessarily the case that this "impossible" matrix will be useless in practice. In genetic applications the sample matrices are used to construct selection indices which are essentially multiple regression predictors of genotype from measurements on one or more traits of an individual and his relatives (e.g., Henderson 1963). A particular difficulty can arise when selection is based on the mean performance of a family, e.g. in a sib or progeny test, and the family size (N) is large (Arnason 1976). The predicted covariance matrix of family means is $\hat{\mathbf{\Psi}}' + \hat{\mathbf{\Sigma}}/N$, and if $\hat{\mathbf{\Psi}}'$ is n.p.d. and N is large, this matrix may be singular or nearly so. Since its inverse is used in the index construction, the computed index becomes very unreliable: at its most extreme the worst rather than the best individuals could be chosen. This is analogous to using a negative

TABLE 4

Approximations for Moments of Sample Roots, $\hat{\alpha}_i$, of **B-W** and Probabilities (Q) of n.p.d. Matrices.

Intra-class correlations (t_i)	0.025 0.0625	0.025 0.125	0.0625 0.125
		s = 160, n = 5	
Roots $\alpha_i = nt_i$	0.125 0.3125	0.125 0.625	0.3125 0.625
Simulated $E(\hat{\alpha}_i)$	0.0664 0.3730	0.1046 0.6524	0.2662 0.6721
Predicted $E(\hat{\alpha}_i)$	0.0713 0.3662	0.1016 0.6484	0.2707 0.6668
Simulated $SD(\hat{\alpha}_i)$	0.1164 0.1350	0.1293 0.1716	0.1325 0.1631
Predicted $SD(\hat{\alpha}_i)$	0.1109 0.1294	0.1305 0.1723	0.1369 0.1655
Simulated Q% ± SE	28.7 ± 1.4	21.8 ± 1.3	2.6 ± 0.5
Predicted Q%	26.0	21.8	2.4
		s = 40, n = 10	
Simulated Q% ± SE	33.4 ± 1.5	26.3 ± 1.4	5.0 ± 0.7
Predicted Q%	29.5	25.0	5.2
		s = 160, n = 10	
Simulated Q% ± SE	4.7 ± 0.7	4.4 ± 0.6	0.0
Predicted Q%	5.3	4.8	0.0

heritability estimate for an index using a single trait. Although in other selection index applications an n.p.d. genetic covariance matrix may be less critical, the problems of using them are being investigated and we intend to report further subsequently. At present, we recommend that investigators check whether the covariance matrix is p.d. and if not, proceed with caution.

Acknowledgments

We are grateful to Marjorie McEwan for computational assistance and to Dr. C. M. Theobald for useful comments.

Résumé

La probabilité (Q) que la matrice de covariance inter-groupe estimée soit non définie positive est calculée pour une analyse de variance multivariate à effets aléatoires pour une seule classification équilibrée. On y montre que Q ne dépend que des valeurs propres de la matrice produit de l'inverse de la vraie matrice de covariance intra-groupes par la matrice de cavriance entre-groupes qui, pur des variables indépendantes s'exprime en fonction des corrélations intra-classes.

Des valeurs de Q sont calculées pour des gammes de tailles d'expériences, de corrélations intra-classes et de nombres de variables. Même pour de grandes expériences, Q peut être proche de 1 s'il y a beaucoup de variables, par exemple avec 160 groupes de taille 10 et 8 variables indépendantes chacune avec un coefficient de corrélation intraclasse de 0,025 ou 14 variables chacune avec un coefficient de 0,0625. Une certaine explication des résultats est donnée en fonction du biais des valeurs propres de la matrice de covariance entre groupes estimée.

Dans les applications génétiques la matrice de covariance entre-groupes est proportionnelle à la matrice de covariance génétique ; si elle est non définie positive les héritabilités et les corrélations génétiques ordinaires ou partielles sont hors de leurs limites de validité et l'impact sur la construction de l'index de sélection est discuté.

References

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis,* Wiley, New York.

Anderson, T. W. (1963). Asymptotic theory for principle component analysis. *Annals of Mathematical Statistics 34,* 122–148.

Anderson, T. W. and Das Gupta, S. (1964). A monotonicity property of the power functions of some tests of the equality of two covariance matrices. *Annals of Mathematical Statistics 35,* 1059–1063.

Arnason, T. (1976). Studies on traits in the Icelandic riding ponies. Unpublished M.Sc. Thesis, University of Edinburgh, Scotland.

Gill, J. L. and Jensen, E. L. (1968). Probability of obtaining negative estimates of heritability. *Biometrics 24,* 517–526.

Graybill, F. A. (1961). *An Introduction to Linear Statistical Models, Volume 1.* McGraw-Hill, New York.

Henderson, C. R. (1963). Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding.* W. D. Hanson and H. F. Robinson (eds.), National Academy of Sciences, National Research Council, Washington, D.C., Publication 982, 141–163.

Lawley, D. N. (1956). Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika 43,* 128–136.

Pillai, K. C. S. and Al-Ani, S. (1970). Power comparisons of tests of equality of two covariance matrices based on individual characteristic roots. *Journal of the American Statistical Association 65,* 438–446.

Seal, H. (1966). *Multivariate Statistical Analysis for Biologists,* Methuen, London.

Smith, C., King, J. W. B. and Gilbert, N. (1962). Genetic parameters of British Large White bacon pigs. *Animal Production 4,* 128–143.

Smith, W. B. and Hocking, R. R. (1972). Algorithm AS53. Wishart variate generator. *Applied Statistics 21,* 341–345.

Venables, W. N. (1974). Algorithm AS77. Null distribution of the largest root statistic. *Applied Statistics 23*, 458–465.

Received March 1977; Revised January 1978

Appendix: Rationalisation of Results

At first sight the high probabilities of n.p.d. matrices are somewhat surprising and in this section we try to give some qualitative justification for them. The formulae involved suggest that a more exact quantitative explanation would be difficult. Q depends on the parameters p, s, n and the vector $\underset{\sim}{\lambda}$ or $\mathbf{t}$, and we first consider the case where the population roots are equal, i.e. $\underset{\sim}{\lambda} = \lambda\mathbf{1}$ so there are just four parameters.

Dependence of Q on n and t for equal roots. When $p = 1$ and $t = 0$ (or $\lambda = 1$), Q can be calculated from F tables and these show that if the denominator d.f., $s(n-1)$, are large, the distribution of F is little affected by changes in $s(n-1)$ or, equivalently, the group size n. Similarly, we might expect the distribution of the smallest root of $\mathbf{F} = \mathbf{W}^{-1}\mathbf{B}$ to be almost independent of n if the d.f. for $\mathbf{W}$ are large. This is in fact so: for example, if $p = 2, s = 20, t = 0$ and $n = 10$ the probability is 0.972 that the smallest root of $\mathbf{F} = \mathbf{W}^{-1}\mathbf{B}$ exceeds 0.5, and the probability is 0.047 that it exceeds 1.1, whereas with $n = 40$ the probabilities are 0.976 and 0.047 of exceeding 0.5 and 1.1, respectively. These values were calculated using Venables algorithm and noting that Prob(largest root of $\mathbf{S}_W(\mathbf{S}_W + \mathbf{S}_B)^{-1} < s(n-1)/[x(s-1) + s(n-1)]$) = Prob(smallest root of $\mathbf{F} < x$). Since the value of $\mathbf{F}$ for general λ is λ times the value for $\lambda = 1$, the probability that the smallest root of $\mathbf{F}$ is less than x for $\lambda = 1$ is equal to the probability that the smallest root of $\mathbf{F}$ for general λ is less than $\lambda x = x[1 + (n-1)t]/(1-t)$, which as n increases and t decreases approaches ntx. This suggests that Q might be approximately independent of n given nt. Examples in Table 1, e.g. $n = 20$, $t = 0.125$ versus $n = 40$, $t = 0.0625$ confirm this.

Mean and variance of smallest root of F for equal roots. A plot of the distribution function of $\mathbf{F}$ for $p = 2$ and $t = 0$ for given n and s suggests that the smallest root of $\mathbf{F}$ is approximately normal with mean $1 - B(p, s, n)$ and variance $[2n/s(n-1)]V(p, s, n)$ when $t = 0$. Because the population value of the smallest root of $\mathbf{F}$ is unity when $t = 0$, $B(p, s, n)$ can be regarded as the bias in the smallest root of $\mathbf{F}$ and $V(p, s, n)$ as the variance in the smallest root, expressed as a proportion of that when $p = 1$ and $t = 0$. Using Venables' algorithm and numerical integration, values for the mean and variance of the smallest root of $\mathbf{F}$ can be calculated. We find both are biased downwards from the univariate case. For example, $B(2, 40, 10) = 0.797$, $B(2, 80, 10) = 0.855$, $B(2, 40, 40) = 0.799$, $V(2, 40, 10) = 0.525$, $V(2, 80, 10) = 0.579$ and $V(2, 40, 40) = 0.537$. This downward bias in the smallest root could be anticipated, since the sum of the roots (equal to the trace of $\mathbf{W}^{-1}\mathbf{B}$) is an unbiased estimate of the sum of the population roots. Again, these are essentially independent of n, so we suggest that the bias and variance functions can be expressed as $B(p, s)$ and $V(p, s)$. The probabilities of n.p.d. matrices can then be approximated by assuming normality of the smallest root, and thus Q would equal 50% if $1/(1 - B(p, s)) = \lambda$.

It is in general difficult to obtain analytic expressions for $B(p, s)$ and $V(p, s)$ since the roots satisfy pth order polynomial equations. If, however, $\mathbf{W}$ is assumed known, the asymptotic results of Anderson (1963) suggest that with equal population roots, $B(2, s)$ is approximately $(\pi/2s)^{1/2}$ and $V(2, s)$ is approximately $(1.5 - \pi/4) = 0.715$. This predicted bias is in excellent agreement with the calculated values but 0.715 overestimates the proportional variance by about $1/\sqrt{s}$ when n is large. Presumably more d.f. are needed for this approximation to be useful. For $p > 2$, Anderson's formulae are difficult to evaluate.

Approximations to Q for unequal roots. If the roots are different and sufficiently large, Lawley's (1956) results using the δ-method for the mean and variance of sample roots of a covariance matrix can be extended to the analysis of variance. It is convenient, for small n, to take the roots, $\hat{\alpha}$, of $\mathbf{B} - \mathbf{W}$ to avoid computing variances of elements of $\mathbf{W}^{-1}$. We obtain

$$E(\hat{\alpha}_i) = nt_i + \sum_j{}' \frac{1}{n(t_i - t_j)}\left\{\frac{[1 + (n-1)t_i][1 + (n-1)t_j]}{s-1} + \frac{(1-t_i)(1-t_j)}{s(n-1)}\right\} + O(s^{-2}), \qquad (A1)$$

where Σ_j' denotes the sum $j = 1, \ldots, p$, excluding $j = i$. This form demonstrates that the terms in brackets in (2) are the variances of B_{ij} and W_{ij}. In terms of the roots, λ_i, and correcting to $0(s^{-2})$

$$E(\hat{\alpha}_i) = nt_i + \sum_j{}' \frac{(1-t_i)(1-t_j)}{sn(t_i - t_j)}\left(\lambda_i\lambda_j + \frac{1}{n-1}\right) + 0(s^{-2}). \qquad (A2)$$

It is clear that the terms due to $V(W_{ij})$ in (3) become relatively small as n increases. Also

$$V(\hat{\alpha}_i) = \frac{2(1-t_i)^2}{n}\left(\frac{\lambda_i^2}{s-1} + \frac{1}{s(n-1)}\right) + 4\sum_j{}' \frac{(1-t_i)^2(1-t_j)}{s^2 n(t_i-t_j)}\left(\lambda_i^2\lambda_j - \frac{1}{(n-1)^2}\right)$$
$$-2\sum_j{}' \frac{(1-t_i)^2(1-t_j)}{s^2n^2(t_i-t_j)^2}\left(\lambda_i\lambda_j + \frac{1}{n-1}\right)\left[2(1-t_i)\left(\lambda_i^2 + \frac{1}{n-1}\right)\right.$$
$$\left. - (1-t_j)\left(\lambda_i\lambda_j + \frac{1}{n-1}\right)\right] + 0(s^{-3}). \qquad \text{(A3)}$$

The formulae also hold for any root which differs from all others, even if some of the latter are the same. Equation (A2) shows that the magnitude of $B(p, s, n)$ or $B(p, s)$ depends on the differences between the smallest and the other roots, and in this case of unequal roots that $B(p, s)$ is proportional to $1/s$, rather than $1/\sqrt{s}$ for the case of equal roots. Equations (A2) and (A3) were used to predict probabilities of n.p.d. matrices by assuming the smallest sample root was normally distributed. Some examples are given in Table 4, to show the usefulness of the approximations both to the moments of the sample roots and to Q.

The above formulae for the unequal roots case also show that $B(p, s)$ increases and $V(p, s)$ decreases linearly with the number of variables, p. Analogy with order statistics, and the roughly normal integral relationship between Q and s shown in Figure 1, suggests that at least for $B(p, s)$ the same approximate linear correspondence with p holds in the equal roots case. Numerical computations generally confirm this argument.

Reprinted from
BIOMETRICS COPYRIGHT © 1978
THE BIOMETRIC SOCIETY, Vol. 34, No. 3, September 1978

23

Bias in variance and covariance component estimators due to selection on a correlated trait

by

K. Meyer and R. Thompson

Sonderdruck aus Zeitschrift für Tierzüchtung und Züchtungsbiologie
Bd. 101 (1984), H. 1, S. 33-50

VERLAG PAUL PAREY · HAMBURG 1 · SPITALERSTRASSE 12

Rechte, auch die der Übersetzung, des Nachdrucks, der photomechanischen Wiedergabe und der Speicherung in Datenverarbeitungsanlagen, vorbehalten, © 1984 Verlag Paul Parey, Hamburg und Berlin

Institute of Animal Genetics and ARC Unit of Statistics, Edinburgh University, Scotland

Bias in variance and covariance component estimators due to selection on a correlated trait

By KARIN MEYER and R. THOMPSON

Ms. received 2. 12. 1982

Introduction

Statistical methods generally used in the analysis of animal breeding data assume that these data are a random sample from the population concerned. However, as livestock improvement programmes consist largely of continuous selection decisions this is often not true, particularly for field data. Analyses neglecting these selection decisions are not likely to give appropriate estimates unless selection were acting on some character independent of the traits under analysis.

Removal of selection bias will normally depend on the correct identification of the selection criteria and is only achieved when all information contributing towards the selection decision is included in the model of analysis. Yet, in many situations this is not the case. For the analysis of dairy records for first and second lactation, for instance, this would require that the decision whether or not a cow was to have a second lactation were solely determined by the first lactation record. In practice, however, additional factors like type, fertility, temperament or health status play an important rôle.

In the simplest case selection might be thought to be on one trait correlated with yield. The objective of this paper is to investigate for this case the consequences of carrying out analyses on yield.

The model

Consider a trivariate normal distribution on three traits c, y_1 and y_2

$$E\begin{pmatrix} c \\ y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \text{ and } \text{Var}\begin{pmatrix} c \\ y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \sigma_c^2 & \sigma_{1c} & \sigma_{2c} \\ \sigma_{1c} & \sigma_1^2 & \sigma_{12} \\ \sigma_{2c} & \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

Let measurements be made for y_1 on all animals but for y_2 on a subset of the sample only. Assume that lack or presence of a record for y_1 is the outcome of a truncation selection on c.

This is similar to ROBERTSON's (1966) model for the culling process amongst dairy cows with c, y_1 and y_2 denoting the culling variate, first and second lactation yield, respectively. The culling variate is a conceptual variable, a kind of selection index combining all factors contributing towards the selection decision, ROBERTSON (1966) partitioned it into two components, first lactation yield and causes of culling other than yield, additionally allowing for random culling.

U.S. Copyright Clearance Center Code Statement: 0044–3581/84/10101–0033 $ 2.50/0
Z. Tierzüchtg. Züchtgsbiol. 101 (1984) 33–50
© 1984 Verlag Paul Parey, Hamburg und Berlin
ISSN 0044–3581/InterCode: ZTZBAS

Means and variances of the multivariate normal distribution after selection have been derived by PEARSON (1903) and presented in matrix notation subsequently (AITKEN 1934; HENDERSON 1975). For truncation selection, the selection differential in c, Δc, is determined by the selection intensity, i_c, ie. $\Delta c = i_c\sigma_c$. Let '*' denote values in the selected population. Then

$$E\begin{pmatrix} c^* \\ y_1^* \\ y_2^* \end{pmatrix} = \begin{pmatrix} \Delta c \\ b_{1.c}\Delta c \\ b_{2.c}\Delta c \end{pmatrix} = \begin{pmatrix} i_c\sigma_c \\ i_c r_{1c}\sigma_c \\ i_c r_{2c}\sigma_c \end{pmatrix}$$

Selection reduces the variance of c by a proportion $K = i_c(i_c - x)$, where x denotes the truncation point of the normal distribution corresponding to the proportion selected, p. For $r_{12} \neq 0$ this gives

$$\mathrm{Var}\begin{pmatrix} c^* \\ y_1^* \\ y_2^* \end{pmatrix} = \begin{pmatrix} (1-K)\sigma_c^2 & \text{symmetric} & \\ (1-K)\sigma_{1c} & (1-Kr_{1c}^2)\sigma_1^2 & \\ (1-K)\sigma_{2c} & (1-Kr_{1c}r_{2c}/r_{12})\sigma_{12} & (1-Kr_{2c}^2)\sigma_2^2 \end{pmatrix} \quad \ldots(2)$$

Ignoring c, the corresponding values under a model assuming selection on y_1 would be

$$E\begin{pmatrix} \tilde{y}_1^* \\ \tilde{y}_2^* \end{pmatrix} = \begin{pmatrix} i_1\sigma_1 \\ i_1 r_{12}\sigma_2 \end{pmatrix} \quad (3) \quad \text{and} \quad \mathrm{Var}\begin{pmatrix} \tilde{y}_1^* \\ \tilde{y}_2^* \end{pmatrix} = \begin{pmatrix} (1-K)\sigma_1^2 & \text{symmetric} \\ (1-K)\sigma_{12} & (1-Kr_{12}^2)\sigma_2^2 \end{pmatrix} \quad \ldots(4)$$

Estimation and prediction of linear effects

Fixed effects

Under a selection model, estimates of fixed effects by ordinary least squares are biassed. LUSH and SHRODE (1950) discussed the bias for different linear models for the example of age effects in dairy cattle. Estimating the difference between first and second lactation assuming selection has been acting on first lactation only, it was shown that the maximum likelihood (ML) or generalised least squares (GLS) estimator is unbiassed (from (3) and (4));

$$E(\widehat{\mu_2 - \mu_1}) = E\left(\tilde{y}_2^* - \tilde{y}_1 + b_{2.1}(\tilde{y}_2^* - \tilde{y}_1^*) - b_{2.1}(\tilde{y}_2^* - \tilde{y}_1)\right) = 0 \qquad \ldots (5)$$

while "gross" $(\tilde{y}_2^* - \tilde{y}_1)$ and "paired" $(\tilde{y}_2^* - \tilde{y}_1^*)$ comparison over – and underestimate, respectively, the true difference.

If selection has been acting on c rather than y_1, however, this estimator can be biassed.

$$E(\widehat{\mu_2 - \mu_1}) = E(y_2^* - b_{2.1}^*(y_1^* - y_1) - y_1) = (b_{2.c} - b_{2.1}^* b_{1.c})\Delta c \qquad \ldots (6)$$

where $b_{2.1}^*$ pinpoints a further potential source of bias – if $b_{2.1}$, the regression of y_2 on y_1 is not known but estimated from the selected data, it will be unbiassed if selection has been acting on y_1 but be biassed downwards for selection on c.

$$b_{2.1}^* = \frac{\sigma_{12}^*}{\sigma_1^2} = \frac{1 - K r_{1c}r_{2c}/r_{12}}{1 - Kr_{1c}^2} b_{2.1} \qquad \ldots (7)$$

Assuming $b_{2.1}$ to be known, the GLS or ML estimator (eq. (6)) is unbiassed even under the wrong model of analysis if $b_{2.c} = b_{2.1}b_{1.c}$ or, equivalently $r_{2c} = r_{12}r_{1c}$.

This holds if the partial correlation of y_2 on c given y_1 is zero, i. e. if all variation in y_2 depending on c is explained by y_1.

Random effects

Similar arguments hold for the prediction of random effects. HENDERSON (1975) examined BLUP under various conceptual selection models, showing that for selection on the vector of observations or linear functions thereof, sire solutions ignoring selection are unbiassed if selection is totally within levels of fixed effects. As HENDERSON's (1975) concepts were formulated in order to meet the requirement of translation invariance of estimators to selection, they are hard to envisage and have as yet not been related to practical data. WICKHAM (1975) derived the probability density function of a bivariate normal distribution after selection on a third unobserved trait and the corresponding ML estimators of means and variances. As these required not only the knowledge of the truncation point and the covariances to the culling variate but also resulted in a set of equations non-linear in all of the parameters, it was concluded that they were of little practical relevance.

Considering records for one sire only, the effect of selection on the GLS predictors of sire effects can be evaluated. Let a sire have $n_1 + n_2$ progeny with records for y_1 with mean $\bar{y}_1$ and expected value S_1, and n_2 progeny with records for y_2 and mean $\bar{y}_2$. Assume no selection apart from involuntary culling which is equally likely to affect all animals has taken place. Means and variances are then the same as if all n_1+n_2 animals had records for y_2 and $E(\bar{y}_2)=S_2$, where S_i denotes the true sire mean for y_i.

Let the general mixed model be

$$y = \underset{\sim}{X}\underline{b} + \underset{\sim}{Z}\underline{u} + \underline{e} \text{ with}$$

$$E\begin{pmatrix} \underline{y} \\ \underline{u} \\ \underline{e} \end{pmatrix} = \begin{pmatrix} \underline{Xb} \\ \underline{0} \\ \underline{0} \end{pmatrix} \quad \text{and Var} \quad \begin{pmatrix} \underline{y} \\ \underline{u} \\ \underline{e} \end{pmatrix} = \begin{pmatrix} \underline{V} & \underline{ZG} & \underline{R} \\ \underline{GZ}' & \underline{G} & \underline{0} \\ \underline{R} & \underline{0} & \underline{R} \end{pmatrix}$$

where $\underline{y}$ denotes the vector of observations
$\underline{b}$ the vector of fixed effects
$\underline{u}$ the vector of random effects,
$\underline{e}$ the vector of residual errors,
$\underline{X}$ the design matrix for fixed effects,
and $\underline{Z}$ the design matrix for random effects.

Sire effects can then be predicted by solving the "Mixed Model Equations" (MME, HENDERSON (1973).

$$\begin{pmatrix} \underline{X}' \underline{R}^{-1} \underline{X} & \underline{X}' \underline{R}^{-1} \underline{Z} \\ \underline{Z}' \underline{R}^{-1} \underline{X} & \underline{Z}' \underline{R}^{-1} \underline{Z} + \underline{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\underline{b}} \\ \hat{\underline{u}} \end{pmatrix} = \begin{pmatrix} \underline{X}' \underline{R}^{-1} \underline{y} \\ \underline{Z}' \underline{R}^{-1} \underline{y} \end{pmatrix}$$

Considering records for one sire only, assuming no fixed effects, this reduces to

$$(\underline{Z}'\underline{R}^{-1}\underline{Z}+\underline{G}^{-1})\,\hat{\underline{u}}=\underline{Z}'\underline{R}^{-1}\underline{y} \qquad \ldots(8)$$

Let $\underline{y}_1$, length n_1, denote the vector of observations for all progeny with only one record and $\underline{y}_2$, length $2n_2$, the corresponding vector of records for y_1 and y_2 for animals having both records,

$$\underline{y} = \begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \end{pmatrix} \quad \text{with} \quad \underline{y}_1 = \left\{ Y_{11i} \right\} \quad i = 1, \ldots, n_1$$

$$\underline{y}_2 = \left\{ \begin{pmatrix} Y_{12i} \\ Y_{22i} \end{pmatrix} \right\} \quad i = 1, \ldots, n_2$$

and partition the design matrix accordingly

$$\underline{Z} = \begin{pmatrix} \underline{Z}_1 \\ \underline{Z}_2 \end{pmatrix} \quad \text{with} \quad \underline{Z}_1 = \left\{ \underline{Z}_{1i} \right\} \quad \text{and } \underline{Z}_{1i} = (1 \quad 0)$$

$$\underline{Z}_2 = \left\{ \underline{Z}_{2i} \right\} \quad \text{and } \underline{Z}_{2i} = \underline{I}_2$$

where I_m denotes an identity matrix of order m. Let $\underset{\sim}{E}$ be the within sire family variance/covariance matrix of y_1 and y_2 and $\underset{\sim}{G}$ be the corresponding matrix for sire effects, u_1 and u_2.

$$\underset{\sim}{E} = \begin{pmatrix} E_{11} & E_{12} \\ E_{12} & E_{22} \end{pmatrix} \quad ; \quad \underset{\sim}{G} = \begin{pmatrix} G_{11} & G_{12} \\ G_{12} & G_{22} \end{pmatrix}$$

$$\text{Then } \underset{\sim}{R} = \begin{pmatrix} \underset{\sim}{R}_1 & 0 \\ 0 & \underset{\sim}{R}_2 \end{pmatrix} \quad \text{with} \quad \begin{aligned} \underset{\sim}{R}_1 &= \underset{\sim}{I}_{n1} \; E_{11} \\ \underset{\sim}{R}_2 &= \underset{\sim}{I}_{n2} * \underset{\sim}{E} \end{aligned}$$

where '*' denotes the direct matrix product (SEARLE 1966).
The MME (8) can then be written as

$$(\underset{\sim}{Z}_1' \underset{\sim}{R}_1^{-1} \underset{\sim}{Z}_1 + \underset{\sim}{Z}_2' \underset{\sim}{R}_2^{-1} \underset{\sim}{Z}_2 + \underset{\sim}{G}^{-1}) \hat{\underset{\sim}{u}} = \underset{\sim}{Z}_1 \underset{\sim}{R}_1^{-1} \underset{\sim}{y}_1 + \underset{\sim}{Z}_2' \underset{\sim}{R}_2^{-1} \underset{\sim}{y}_2 \qquad \ldots (9)$$

with $\hat{\underset{\sim}{u}}' = (\hat{u}_1 \; \hat{u}_2)$ this gives

$$\left[n_1 \begin{pmatrix} E_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} + n_2 \begin{pmatrix} E_{11} & E_{12} \\ E_{12} & E_{22} \end{pmatrix}^{-1} + \begin{pmatrix} G_{11} & G_{12} \\ G_{12} & G_{22} \end{pmatrix}^{-1} \right] \begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \end{pmatrix}$$

$$= n_1 \begin{pmatrix} E_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \bar{y}_{11} \\ 0 \end{pmatrix} + n_2 \begin{pmatrix} E_{11} & E_{12} \\ E_{12} & E_{22} \end{pmatrix}^{-1} \begin{pmatrix} \bar{y}_{12} \\ \bar{y}_2 \end{pmatrix} \qquad \ldots (10)$$

where $\bar{y}_{12}$ denotes the mean of y_1 for all animals having both records and $\bar{y}_{11}$ the corresponding figure for animals having only one record.

From $\bar{y}_1 = \dfrac{n_1 \bar{y}_{11} + n_2 \bar{y}_{12}}{n_1 + n_2}$ it follows that $\bar{y}_{11} = \bar{y}_1 - \dfrac{n_2}{n_1} (\bar{y}_{12} - \bar{y}_1)$

where $\bar{y}_{12} - \bar{y}_1$, of course, is the selection differential in y_1. In the no selection case $\bar{y}_1 = \bar{y}_{11} = \bar{y}_{12}$. Replacing the sire means by their expectations and simplifying then reduces the RHS, $\underset{\sim}{r}$ of the MME (10) to

$$\underset{\sim}{r} = \frac{n_2}{E_{11}E_{22} - E_{12}^2} \begin{pmatrix} E_{22} S_1 - E_{12} S_2 + \dfrac{n_1}{n_2} (E_{22} - E_{12}^2/E_{11}) S_1 \\ E_{11} S_2 - E_{12} S_1 \end{pmatrix},$$

Assume now, selection on c has been carried out with a response of $\Delta c = i_{wc} \sigma_{wc}$, where i_{wc} denotes the within sire family selection intensity and σ^2_{wc} the corresponding within family variance. The expectations of sire means after selection are then

$$\begin{aligned} E(\bar{y}_1) &= S_1 \\ E(\bar{y}_{12}) &= S_1 + b_{w1.c} \Delta c \\ E(\bar{y}_{11}) &= S_1 - \frac{n_2}{n_1} b_{w1.c} \Delta c \\ E(\bar{y}_2) &= S_2 + b_{w2.c} \Delta c \end{aligned} \qquad \ldots (11)$$

where $b_{wi.j}$ denotes the within family regression coefficient of i on j.
Again, substituting the sire means by their expectations gives the expected RHS after selection as

$$\underset{\sim}{r}^{*} = \frac{n_2}{E_{11}E_{22} - E_{12}^2} \begin{pmatrix} E_{22}\, S_1 - E_{12}\, S_2^{*} + \frac{n_1}{n_2}\,(E_{22} - E_{12}^2/E_{11})S_1 \\ E_{11}\, S_2^{*} - E_{12}\, S_1 \end{pmatrix}$$

$$\text{with} \quad S_2^{*} = S_2 + (b_{w2.c} - b_{w2.1}\, b_{w1.c})\Delta c \qquad \ldots (12)$$

i. e. selection replaces S_2 by S_2^{*} while S_1 remains unaltered.

Thus under a selection model unbiassed sire solutions will be obtained if $b_{w2c} = b_{w2.1}b_{w1.c}$. This is the case when $c = y_1$, alternatively when $r_{w2c.1}$, the within family partial correlation between c and y_2 given y_1 is zero. The latter is the same condition as for the unbiassed estimation of fixed effects, but on the within family rather than the phenotypic level.

Estimation of variance components

Within family

Consider an individual sire family as a "population" with means S_{ci}, S_{1i} and S_{2i}, variances, covariances σ^2_{wc}, σ_{w2c}, σ^2_{w1}, σ_{w12} and σ^2_{w2}. Application of the Pearson results gives the expected means after selection as

$$\begin{aligned} E(S_{ci}^{*}) &= S_{ci} + i_{wi}\, \sigma_{wc} \\ E(S_{1i}^{*}) &= S_{1i} + b_{w1.c}\, i_{wi}\, \sigma_{wc} \\ E(S_{2i}^{*}) &= S_{2i} + b_{w2.c}\, i_{wi}\, \sigma_{wc} \end{aligned} \qquad \ldots (13)$$

Analogously, variance/covariance components affected by selection are

$$E(\sigma_{w12}^{*}) = \sigma_{w12} - K_i \frac{\sigma_{w1c}\,\sigma_{w2c}}{\sigma^2_{w2}} = \left(1 - K_i \frac{r_{w1c}\, r_{w2c}}{r_{w12}}\right)\sigma_{w12}$$

$$\text{for } r_{w12} \neq 0 \qquad \ldots (14)$$

$$E(\sigma_{w2}^{*2}) = \sigma^2_{w2} - K_i \frac{\sigma^2_{w2c}}{\sigma^2_{wc}} = (1 - K_i\, r^2_{w2c})\sigma^2_{w2} \qquad \ldots (15)$$

with $K_i = i_{wi}(i_{wi} - x_i)$ where i_{wi} denotes the selection intensity within the ith sire family and x_i the corresponding truncation point expressed in standard units.
AOV: The analysis of variance/covariance (AOV) will use these components, σ^{*}_{w12} and σ^{2*}_{w2}, pooled over all sire families to estimate the within family mean squares. Approximation of the average of K_i by K_o, the value of K_i for a sire family with mean equal to the population mean, gives the expectations of AOV estimates as

$$\begin{aligned} E(\sigma^2_{w1}) &= \sigma^2_{w1} \\ E(\sigma_{w12}) &= \left(1 - K_o \frac{r_{w1c}\, r_{w2c}}{r_{w12}}\right)\sigma_{w12} \quad (\text{for } r_{w12} \neq 0) \\ E(\sigma^2_{w2}) &= (1 - K_o\, r^2_{w2c})\, \sigma^2_{w2} \end{aligned} \qquad \ldots (16)$$

ML: Maximum likelihood (ML) estimation of within family components makes use of two properties of the multivariate normal distribution. Firstly, the regression of an indirectly on the directly selected variable(s) is not affected by selection. Secondly, the variance about regression, i. e. the conditional variance of the trait under indirect selection remains unchanged (e. g. CURNOW 1961). Treating y_1 as apparent selection criteria, ML will estimate

$$\hat{\sigma}_{w12} = \frac{\sigma^*_{w12}}{\sigma^{2\,*}_{w1}} \sigma^2_{w1} \qquad \ldots (17)$$

with selection in fact acting on c, however, this has expectation

$$E(\hat{\sigma}_{w12}) = \frac{1 - K_i \frac{r_{w1c} r_{w2c}}{r_{w12}}}{1 - K_i r^2_{w1c}} \sigma_{w12} \qquad \ldots (18)$$

i. e. is on average biased by

$$\frac{K_o r_{w1c} (r_{w1c} - r_{w2c}/r_{w12})}{1 - K_o r^2_{w1c}} \sigma_{w12} \qquad \ldots (19)$$

Hence, σ_{w12} is unbiassed if $r_{w1c} = 0$ or $r_{w2c} = r_{w12} r_{w1c}$, i. e. $r_{w2c.1} = 0$.
Correspondingly, ML will estimate

$$\hat{\sigma}^2_{w2} = \sigma^{2*}_{w2} - \frac{(\sigma^*_{w12})^2}{\sigma^{2\,*}_{w1}} + \frac{\hat{\sigma}^2_{w12}}{\sigma^2_{w1}} \qquad \ldots (20)$$

which has expectation

$$E(\hat{\sigma}^2_{w2}) = \left\{1 - K_i r^2_{w2c} + K_i r^2_{w1c} r^2_{w12} \left(\frac{1 - K_i r_{w1c} r_{w2c}/r_{w12}}{1 - K_i r^2_{w1c}}\right)^2\right\} \sigma^2_{w2} \qquad \ldots (21)$$

and is on average biassed by

$$-K_o \frac{[r_{w2c} + r_{w1c} (r_{w12} - 2K_o r_{w1c} r_{w2c})]}{(1 - K_o r^2_{w1c})^2} (r_{w2c} - r_{w1c} r_{w12}) \sigma^2_{w2} \qquad \ldots (22)$$

i. e., again, is unbiassed if $r_{w2c.1} = 0$.

Sire components

The within sire family selection intensity, i_{wi}, is a function of the mean culling variate, S_{ci}. Using a Taylor series expansion to obtain a linear prediction of i_{wi} gives (evaluated at $S_{ci} = S_{co}$)

$$i_{wi} = i_{wo} - \frac{K_o}{\sigma_{wc}} S_{ci} \qquad \ldots (23)$$

Hence, from (13),

$$\begin{aligned} E(S^*_{1i}) &= S_{1i} + b_{w1.c}\, i_{wo}\, \sigma_{wc} - K_o b_{w1.c}\, S_{ci} \\ E(S^*_{2i}) &= S_{2i} + b_{w2.c}\, i_{wo}\, \sigma_{wc} - K_o b_{w2.c}\, S_{ci} \end{aligned} \qquad \ldots (24)$$

AOV: The AOV will estimate the sire variance/covariance components σ_{Si2}, from

$\hat{\sigma}_{S12} = \mathrm{Cov}(S^*_{1i}, S^*_{2i})$ which has expectation

$$E(\hat{\sigma}_{S12}) = \sigma_{S12} - K_o(b_{w2.c}\,\sigma_{S1c} + b_{w1.c}\,\sigma_{S2c}) + K_o^2\, b_{w1.c}\, b_{w2.c}\,\sigma^2_{Sc}$$

$$= \left\{1 - K_o\left(b_{w2.c}\frac{b_{Sc.1}}{b_{S2.1}} + b_{w1.c}\frac{b_{Sc.2}}{b_{S1.2}}\right) + K_o^2\, b_{w1.c}\, b_{w2.c}\right.$$

$$\left.\frac{b_{Sc.1}}{b_{S1.c}\,b_{S2.1}}\quad \sigma_{S12}\right\} \qquad \ldots (25)$$

where b_S denotes the genetic regression coefficient.

Correspondingly, $\hat{\sigma}^2_S = V(S^*_{2i})$ with

$$E(\hat{\sigma}^2_{S2}) = \sigma^2_{S2} - 2K_o\, b_{w2.c}\,\sigma_{S2c} + K_o^2\, b^2_{w2.c}\quad \sigma^2_{Sc}$$

$$= \left\{1 - 2K_o\, b_{w2.c}\, b_{Sc.2} + K_o^2\, b^2_{w2.c}\quad \frac{b_{Sc.2}}{b_{S2.c}}\right\}\sigma^2_{S2} \qquad \ldots (26)$$

ROBERTSON (1966) has given expressions of the same form, but not exactly the same. To relate the two approaches, (26) can be derived by arguments similiar to those of Robertson. For a sire with mean genetic value of S_{ci}, the mean of survivors will be $E(S^*_{ci}) = S_{ci} + i_{wi}\,\sigma_{wc}$ and the expected value of S_{2i} given selection

$$E(S^*_{2i}|S_{ci}) = b_{S2c}\, S_{ci} + b_{w2.c}\, i_{wi}\,\sigma_{wc}\quad (b_{S2.c} - K_o b_{w2c})$$

$$S_{ci} + b_{w2.c}\, i_{wo}\,\sigma_{wo} \qquad \ldots (27)$$

The conditional variance of S_{2i} given S_{ci} is not affected by selection

$$V(S_{2i}|S_{ci}) = \sigma^2_{S2} - \frac{\sigma^2_{S1c}}{\sigma^2_{Sc}} \qquad \ldots (28)$$

The variance of S_{2i} given selection on c can then be thought of as the sum of these two components,

$$V(S^*_{2i}) = V(S_{2i}|S_{ci}) + V\left(E(S^*_{2i}|S_{ci})\right) \text{ which simplifies to (26).}$$

ML: Estimates of sire variance/covariance components can be obtained using quadratic functions of the sire solution from the mixed model (PATTERSON, THOMPSON 1971; HENDERSON 1973; HENDERSON and QUAAS 1977). As shown selection will alter the MME by replacing S_{2i} in the RHS by $S^*_{2i} = S_{2i} + (b_{w2.c} - b_{w2.1}b_{w1.c})\, i_{wi}\sigma_{wc}$.
Thus ML will estimate

$\hat{\sigma}_{S12} = \mathrm{Cov}(S_{1i}, S^*_{2i})$ and

$\hat{\sigma}^2_{S2} = V(S^*_{2i})$

Replacing i_{wi} by its linear approximation (23) gives

$$E(\hat{\sigma}_{S12}) = [1 - K_o(b_{w2.c} - b_{w2.1}\, b_{w1.c})\quad \frac{b_{Sc.1}}{b_{S1.c}}]\,\sigma_{S12} \qquad \ldots (29)$$

$$E(\hat{\sigma}^2_{S2}) = [1 - 2K_o(b_{w2.c} - b_{w2.1}\, b_{w1.c})b_{Sc.2} + K_o^2\,(b_{w2.c} - b_{w1.c}b_{w2.1})^2 \frac{b_{Sc.2}}{b_{S2.c}}\,]\,\sigma^2_{S2} \quad \ldots (30)$$

Let $q^2 = \frac{\lambda_2}{\lambda_c}$ with $\lambda_i = \frac{\sigma^2_{wi}}{\sigma^2_{Sci}}$. The expected bias in these components is then

$$\left\{-K_o\ \frac{r_{g1c}}{r_{g12}}\,(r_{w2c} - r_{w12}\ r_{w1c})q\right\}\sigma_{S12} \quad \ldots (31)$$

for $\hat{\sigma}_{S12}$ and

$$\left\{-2K_o\ r_{g2C}\,(r_{w2c} - r_{w12}r_{w1c})q + [K_o(r_{w2c} - r_{w1c})q]^2\right\}\sigma^2_{S2} \quad \ldots (32)$$

for $\hat{\sigma}^2_{w2}$, respectively.

Extending the model of analysis

The culling variate, c, can be regarded as the underlying continous distribution for the all-or-none trait survival. Though c is not measurable, the latter can be observed as lack or presence of a record for y_2. Including survival as additional trait in the analysis will reveal some information about the culling variate.

Linear effects

The effect of selection on estimates of fixed and prediction of random effects can be evaluated by a natural extension of the preceding formulae to the multivariate situation. Let x denote the discrete variable survival with x = 0 if an animal has no record for y_2 and x = 1, if both y_1 and y_2 are measured with x and y_1 being the apparent selection criteria. An obvious model to fit is then

$$y_2^* = y_2 + b_{21.x}(y_1^* - y_1) + b_{2x.1}(x^* - x) + e \quad \ldots (33)$$

where $b_{1i.K}$ denotes the partial regression coefficient of i on j given k. In the regular case these coefficients will be estimated using variances and covariances amongst survivors and will be unbiased if all information contributing to the selection decision is included in the model. With x being an all-or-none trait, however, it is not possible to seperate y_2 and $b_{2x.1}$ from within sire family data as each survivor has $x^* = 1$. Hence only $y_2 + b_{2x.1}\,(1-x)$ is estimable. In the following it will be assumed that $b_{2x.1} = 0$, so that y_2 in fact estimates $y_2 + b_{2x.1}\,(1-x)$ and, similiarly, $b_{21.x}$ will be estimated by $b_{2.1}$. This implies that including x into the analysis has not benefit towards the removal of selection bias – equations for fixed and random effects solutions reduce to those from the two-trait model with y_1 and y_2 only.

Estimation of variance components within family

Let p_i denote the proportion of survivors in the ith sire family, q_i the corresponding proportion culled.

$$p_i = \frac{n_{2i}}{n_{1i} + n_{2i}}, \quad q_i = \frac{n_{1i}}{n_{1i} + n_{2i}} = 1 - p_i$$

Further, let $\bar{x}_i$, $\bar{x}_{1i}$, $\bar{x}_{2i}$ denote the ith sire mean for x of all progeny, non-survivors and survivors respectively. It follows that

$$\begin{aligned} E(\bar{x}_i) &= S_{xi} = p_i \\ E(\bar{x}_{1i}) &= S_{x1i} = 0 \\ E(\bar{x}_{2i}) &= S_{x2i} = 1 \end{aligned} \quad \ldots (34)$$

regardless whether culling has been voluntary or by chance. This gives the variance/covariance components involving x as

$$\begin{aligned} \sigma^2_{wx} &= q_i S^2_{x1i} + p_i S^2_{x1i} + p_i S^2_{x2i} - S^2_{xi} = p_i q_i \\ \sigma_{w1x} &= q_i S_{x1i} S_{11i} + p_i S_{x2i} S_{12i} - S_{xi} S_{1i} = p_i S_{12i} \\ \sigma_{w2x} &= p_i S_{x2i} S_{2i} - p_i S_{x2i} S_{2i} = 0 \end{aligned} \quad \ldots (35)$$

Using (1) and (12) gives $\sigma_{w1x} = \frac{z_i}{\sigma w_c} \sigma_{w1c}$,

i. e. the covariance on the underlying scale transformed by a factor depending on the truncation point and the variance of the culling variate. The covariance between x and y_2 is zero as each animal having a record for y_2 has a value of x = 1.

AOV: The AOV will use these components pooled over all sire families to estimate the within family components. Using a linear approximation for p_i as before for K_i gives expected estimates

$$\sigma^2_x = p_o q_o \text{ and } \sigma_{w1x} = (z_o/\sigma_{wc})\sigma_{w1c}$$

where p_o and Z_o denote p_i and Z_i at $S_{ci} = 0$ respectively. Estimates of σ^2_{w1}, σ_{w12} and σ^2_{w2} are not affected by including x.

ML: Extending the framework to a multiple regression situation, ML will estimate

$$\hat{\sigma}_{w2x} = b_{w2x.1} \hat{\sigma}^2_{wx} + b_{w21.x} \hat{\sigma}_{w1x} \quad \ldots (35)$$

$$\hat{\sigma}_{w12} = b_{w2x.1} \hat{\sigma}_{w1x} + b_{w21.x} \hat{\sigma}^2_{w1} \quad \text{and} \quad \ldots (36)$$

$$\hat{\sigma}^2_{w2} = \sigma^{2*}_{w2} - b_{w2x.1}(\sigma^*_{w2x} - \hat{\sigma}_{w2x}) - b_{w21.x}(\sigma^*_{w12} - \hat{\sigma}_{w12}) \quad \ldots (37)$$

As explained for the corresponding regression coefficients on the phenotypic level, $b_{w2x.1} = 0$ and $b_{w21.x} = b_{w2.1}$ if x is an all-or-none trait. Thus (36) and (37) reduce to (17) and (20), i. e. estimates will be the same as in the basic model. The ML procedure will also give a figure for σ_{w2x} which, however, does not provide any extra information allowing to disentangle r_{w2c} and r_{w12}.

$$\hat{\sigma}_{w2x} = \frac{\sigma^*_{w12}}{\sigma^{2*}_{w1}} \hat{\sigma}_{w1x} \quad \ldots (38)$$

$$\text{with } E(\hat{\sigma}_{w2x}) = \frac{1 - Ko\, r_{w1c} r_{w2c} / r_{w12}}{1 - Ko\, r^2_{w1c}} Zo\, r_{w1c} r_{w12} \sigma_{w2} \quad \ldots (39)$$

Sire components

Using a Taylor series gives a linear approximation of the proportion selected with the ith sire family of

$$p_i = p_o + \frac{Z_o}{\sigma_{wc}} \; s_{ci} \qquad \ldots \quad (40)$$

Hence, variances/covariances involving x will be estimated as

$$\hat{\sigma}^2_{Sx} = V(p_i) \text{ with } E(\hat{\sigma}^2_{Sx}) = \frac{Z_o^2}{\sigma^2_{wc}} \; \sigma^2_{Sc} \qquad \ldots \quad (41)$$

$$\text{and } \hat{\sigma}_{S1x} = \text{Cov}(p_{i1} s_{1i}) \text{ with } E(\hat{\sigma}_{S1x}) = \frac{Z_o}{\sigma_{wc}} \; \sigma_{S1c} \qquad \ldots \quad (42)$$

While (41) and (42) hold for both AOV and ML the two methods differ in estimating σ_{S2x}. An AOV using records of survivors only would find no environmental covariance between x and y_2 as $S_{x2i} = 1$ for all sire families. Alternatively, using records for all animals, the AOV would estimate σ_{S2x} subject to selection bias as

$$\hat{\sigma}_{S2x} = \text{Cov}\,(p_{i1}, s^*_{2i}) \text{ with } E(\sigma_{S2x}) = \frac{Z_o}{\sigma_{wc}} \left(1 - K_o \frac{b_{w2.c}}{b_{S2.c}}\right) \sigma_{S2c} \qquad \ldots \quad (43)$$

(s^*_{2i} from (24))

The ML will determine σ_{S2x} as

$$\hat{\sigma}_{S2x} = \text{Cov}\,(p_i, s^*_{2i}) \text{ with } s^*_{2i} \text{ from (12)}$$

$$\text{and } E(\sigma_{S2x}) = \frac{Z_o}{\sigma_{wc}} \left(1 - K_o \frac{b_{w2.c} - b_{w21} b_{w1.c}}{b_{S2c}}\right) \sigma_{S2c} \qquad \ldots \quad (44)$$

Again, estimates of σ^2_{S1}, σ_{s12} and σ^2_{S2} are as in the basic model.

Genetic parameters

The following genetic parameters for the culling variate can be determined:

$$\hat{h}^2_x = \frac{4 . \sigma^2_{Sx}}{\sigma^2_{px}} \text{ with } E(\hat{h}^2_x) = 4\frac{Z_o^2}{pq} \; \frac{\sigma^2_{Sx}}{\sigma^2_{w_c}} \qquad \ldots \quad (45)$$

Replacing Z_o/σ_{wc} by Z/σ_{pc} then yields $E(\hat{h}^2_x) = \frac{Z^2}{pq} \; h^2_c$

(ROBERTSON and LERNER 1949). As p can be observed and Z and q determined accordingly, h^2_c can be estimated as

$$\hat{h}^2_c = \frac{pq}{Z^2} \; \hat{h}^2_x \qquad \ldots \quad (46)$$

$$\hat{r}_{w2x} = \frac{\hat{\sigma}_{w1x}}{\sqrt{\hat{\sigma}^2_{wx} \hat{\sigma}^2_{w1}}} \text{ with } E(\hat{r}_{w1x}) = \sqrt{\frac{Z_o^2}{p_o q_o}} \; r_{w1c} \qquad \ldots \quad (47)$$

From p, the overall proportion surviving, the truncation point in standard units, t, can be determined. Let t_o denote the corresponding within family value for $S_{ci} = 0$. Then

$$t \, \sigma_{pc} = t_o \, \sigma_{wc} \text{ and } t_o = \frac{t}{\sqrt{1 - \frac{1}{4} h^2_c}} \qquad \ldots \quad (48)$$

Having estimated h_c^2, t_o, p_o, q_o and Z_o can be found and r_{w1c} be estimated as

$$\hat{r}_{w1c} = \sqrt{\frac{p_o q_o}{Z_o^2}}\, r_{w1x} \qquad \ldots (49)$$

$$\hat{r}_{g1x} = \frac{\hat{\sigma}_{S1x}}{\sqrt{\hat{\sigma}^2_{Sx}\hat{\sigma}^2_{S1}}} \quad \text{with } E(\hat{r}_{g1x}) = r_{g1c} \qquad \ldots (50)$$

i. e. the genetic correlation between a quantitative and an all-or-none trait is estimated as if the latter were measured on the underlying scale. Simulation results in the literature support this result (OLAUSSON and RØNNINGEN 1975; COX 1974)

$$\hat{r}_{g2x} = \frac{\hat{\sigma}^*_{S2x}}{\sqrt{\hat{\sigma}^2_{Sx}\hat{\sigma}^{2\,*}_{S2}}} \quad \text{with } E(\hat{r}_{g2x}) = r^*_{g2c} \qquad \ldots (51)$$

Similarly to r_{g1x} this correlation is estimated free of scale effects, however, is subject to selection bias.

Simulation study

A simulation study was conducted in order to evaluate the effect of selection on the estimation of variance components empirically. Data were generated for s sire families of size n using pseudo-random numbers. Upon specification of variances, heritabilities and correlations, sire and error effects were sampled from the normal distribution, resulting in a trivariate normal distribution corresponding to c, y_1 and y_2. Simultaneously the true variances and covariances in the sample were determined. Truncation selection was carried out alternatively taking c and y_1 as selection criteria. Culling rates investigated were 30% and 50%, corresponding to truncation points of $-0.5\,\sigma_p$ and 0, respectively. Records for y_1 and y_2 were then analysed by both an analysis of variance/covariance for an unbalanced one-way classification (AOV) and the corresponding maximum likelihood procedure (ML).

The ML procedure used was based on ANDERSON's (1973) algorithm. As described by HARVILLE (1977), this method makes use of the linearity of the variance/covariance matrix of the vector of observations, y, in the parameters to be estimated, $\Theta = \{\Theta_i\}$. Setting the partial derivatives of the log likelihood of $\underset{\sim}{y}$, log L, with respect to the Θ_i to zero is equivalent to equating quadratic functions of the sire and error effects to their expectations. This yields a set of equations linear in the parameters, Θ_i, with the expectations of the second derivatives of log L as coefficients. Hence, applying successive approximations results in an iterative solution scheme by Fisher's method of scoring. With the correct variances/covariances as starting values this procedure usually converged in two to five rounds of iteration.

Table 1 shows estimates of variance/covariance components when the correct model of analysis is used. Without selection, i. e. balanced design, both methods give identical estimates of the within family comonents. The ML estimates of sire components, however, are biassed downwards as ML does not account for the loss in degrees of freedom due to estimating fixed effects. Using a restricted maximum likelihood (REML) procedure would have accounted for this bias but increased the computational requirements considerably and was therefore disregarded. In the balanced one-way classification the following relationship holds (using results from CORBEIL and SEARLE 1976):

$$(\hat{\sigma}_{sij})_{AOV} = \frac{s}{s-1}\,(\hat{\sigma}_{sij})_{ML} + \frac{1}{n(s-1)}\,\hat{\sigma}_{wij}$$

This bias, however, will be ignored in the following and ML estimates called unbiassed if they coincide with the corresponding estimate in the no selection case.

Table 1. **Effect of selection on y_1 on estimates of variance and covariance by analysis of variance (AOV) and maximum likelihood (ML).**
Means over 10 replicates, 150 sires with 30 daughters each, culling rate 30%

	Population value	Sample	No selection AOV	s.e.[x]	Selection on y_1 AOV	s.e.	No selection ML	s.e.	Selection on y_1 ML	s.e.
σ^2_{s1}	6.25	6.36	6.47	0.23	6.47	0.23	6.41	0.23	6.41	0.23
σ_{s12}	5.00	5.12	5.36	0.20	1.75	0.12	5.32	0.20	5.16	0.18
σ^2_{s2}	6.25	6.42	6.68	0.31	4.19	0.25	6.62	0.31	6.29	0.28
σ^2_{w1}	93.75	94.23	94.21	0.61	94.21	0.61	94.21	0.61	94.21	0.61
σ_{w2}	45.00	45.10	45.14	0.47	23.00	0.40	45.14	0.47	44.68	0.84
σ^2_{w2}	93.75	93.73	93.82	0.69	83.15	0.54	93.82	0.69	93.25	0.73

Theoretical values for genetic parameters : all h^2 = 0.25
all σ^2_p = 100
rg_{12} = 0.8
rp_{12} = 0.5
rw_{12} = 0.48

x) Empirical standard error of the mean over replicates

While AOV estimates, covariances in particular, are severely biassed, ML estimates do not seem to be affected by selection. Though estimates in this example are slightly reduced compared to those before selection, this is well within the range of sampling errors and in numerous other sets of simulated data has been found to be due to chance and not representing any selection bias. These results agree with work by ROTHSCHILD et al. (1979) who, in comparing HENDERSON's method 1 and an ML procedure for a selection rate of 50%, reported severely biassed estimates for the former but estimates close to the true values for the latter method.

When data are selected on the basis of c, c not being included in the analysis, ML estimates, however, are biassed by selection (Table 2). In this case AOV estimates are less affected as for selection on y_1 but still substantially more biassed than the corresponding ML

Table 2. **Effect of selection on c on estimates of variance/covariance components for AOV, ML and ML including an all-or-none trait for survival.**
Means over 10 replicates, 150 sires with 30 daughters each, culling rate 30%

	Sample	No selection AOV	s.e.	Selection on c AOV	s.e.	No selection ML	s.e.	Selection on c ML	s.e.	Selection on c ML (incl. 0/1)	s.e.
σ^2_{s1}	6.36	6.47	0.23	6.47	0.23	6.41	0.23	6.41	0.23	6.41	0.23
σ_{s12}	5.12	5.36	0.20	3.38	0.14	5.32	0.20	4.57	0.22	4.51	0.22
σ^2_{s2}	6.42	6.68	0.31	4.54	0.24	6.62	0.31	5.19	0.28	5.10	0.28
σ^2_{w1}	94.23	94.21	0.61	94.21	0.61	94.21	0.61	94.21	0.61	94.21	0.61
σ_{w12}	45.10	45.14	0.47	36.21	0.59	45.14	0.47	39.30	0.67	38.82	0.64
σ^2_{w12}	93.73	93.82	0.69	83.33	0.76	93.82	0.69	84.61	0.81	84.23	0.79

Theoretical values for genetic parameters : all h^2 = 0.25
all σ^2_p = 100
$r_{w2c.1}$ = 0.33

	1c	2c	12
r_g	0.70	0.80	0.80
r_p	0.45	0.49	0.50
r_w	0.43	0.47	0.48

estimates. SCHAEFFER and SOONG (1978) conducted a similar simulation study, involving estimation of variance/covariance components for two traits with zero error covariance under a selection model by REML. In accordance with the present results, estimates for all components were found to be unbiassed if a multi-trait analysis including the selection criteria was carried out but biassed otherwise. Table 2 also illustrates that an extension of the analysis to include survival as an all-or-none trait does not remove selection bias – apart from sampling errors estimates are identical to those from the analysis including y_1 and y_2 only.

Simulation was used to check the validity of the prediction formula for the selection bias empirically. Table 3 gives for a large set of data AOV estimates before and after selection and their predictions. Estimates in the selection situation were predicted using the true values of genetic parameters in the sample and show excellent agreement with the observed values, considering that expected standard errors in the sample would be of the order of 0.4 to 0.5 for all variance components.

Table 3. **AOV estimates for selection on c and predicted values.**
Culling rate 30%, 300 sires with 20 progeny each; one replicate

	Population value	Sample	No sel.	Sel. on c	Predict
σ_{s12}	10.00	9.85	9 79	5.54	5.66
σ^2_{s2}	12.50	12.15	12.21	7.88	7.98
σ_{w12}	40.00	40.27	40.42	31.82	31.09
σ^2_{w2}	87.50	87.52	87.73	79.16	78.43

Table 4. **ML estimates and predicted estimates for selection on c.**
Means over 25 replicates; culling rate 30%, 250 sires with 40 progeny each

	Population value	Sample	Selection on c	Predict
σ_{s12}	5.00	4.98 $\pm$ .08	4.19 $\pm$ .11	4.43
σ^2_{s2}	6.25	6.15 $\pm$ .10	4.85 $\pm$ .13	5.16
σ_{w12}	45.00	44.89 $\pm$ .25	38.26 $\pm$ .24	38.05
σ^2_{w2}	93.75	93.69 $\pm$ .29	84.40 $\pm$ .32	84.03
h^2_2	0.250	0.246 $\pm$ .004	0.218 $\pm$ .006	0.232
r_{gl2}	0.800	0.802 $\pm$ .004	0.770 $\pm$ .009	0.778

Table 4 contrasts means of ML estimates over 25 replicates with their predicted values. Again, predictions are very close to the estimates, particularly for the within family components. Table 5 gives AOV and ML estimates for the same set of data at a culling rate of 50%. Again there is a good fit between observed and predicted values and numerous simulation runs for other constellations of genetic parameters gave corresponding results. It can therefore be concluded that the prediction formulae (16), (18), (21), (25), (26), (29) and (30) give a valid description of the effects of selection on estimates of variance/covariance components by AOV and ML.

ML estimates of genetic parameters for the culling variate are presented in Table 6. Transforming estimates from an analysis including x instead of c gives for parameters not affected by selection values slightly lower than the sample values, however, figures still show

reasonable agreement. Such an analysis, incorporating the appropriate scale transformations, is therefore expected to provide unbiassed estimates of the heritability of the culling variate and its genetic and phenotypic correlation to y_1. The genetic correlation to y_2 will be underestimated due to selection.

Table 5. **Effect of selection with a culling rate of 50% on AOV and ML estimates of variance/covariance components and predicted estimates.**
Means over 50 replicates, 40 sires with 100 daughters

				Analysis of variance					Maximum likelihood[xx]				
	Population value	Sample Mean	Sample s.e.[x]	No selection Mean	No selection s.e.	Selection on c Mean	Selection on c s.e.	Predict	No selection Mean	No selection s.e.	Selection on c Mean	Selection on c s.e.	Predict
σ^2_{s1}	6.25	6.00	.20	5.98	.23	5.98	.23	6.00	5.81	.22	5.81	.22	6.00
σ_{s12}	5.00	4.72	.18	4.73	.21	2.86	.18	2.81	4.60	.20	3.85	.19	4.03
σ^2_{s2}	6.25	6.09	.19	6.13	.21	4.44	.19	4.29	5.96	.20	4.83	.19	5.01
σ^2_{w1}	93.75	92.24	.37	93.42	.32	93.42	.32	92.24	93.42	.31	93.42	.32	92.24
σ_{w12}	45.00	44.21	.32	44.91	.25	31.09	.25	30.41	44.91	.25	36.24	.28	35.65
σ^2_{w2}	93.75	93.07	.35	93.65	.32	79.03	.29	78.99	93.65	.32	81.06	.31	81.02

Theoretical values for genetic parameters :

		1c	2c	12
all h^2 = 0.25	r_g	0.70	0.65	0.80
all σ^2_p = 100	r_p	0.50	0.50	0.50

x) Empirical standard error of means over replicates

xx) Iteration until changes $\leq 2\%$

$r_{w2c.1} = 0.335$

Table 6. **ML estimates of parameters for the culling variate.**
Means over 25 replicates, 100 sires with 40 progeny each, culling rate 30%

		Sample		ML including c		ML including survival (0/1) Observed		ML including survival (0/1) Transformed	
	Population	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
σ^2_{sc}	12.50	11.95	.36	11.78	.34	0.014	.001	10.44	.73
σ_{s1c}	8.75	8.20	.28	8.15	.31	0.286	.012	7.72	.32
σ_{s2c}	8.125	7.70	.30	7.87	.29	0.213	.010	5.75	.27
σ^2_{wc}	87.50	87.92	.41	88.06	.41	0.200	.001		
σ_{w1c}	41.25	41.57	.31	41.55	.31	1.466	.016	39.60	.43
σ_{w2c}	41.875	41.92	.30	41.86	.36	0.496	.007		
h^2_c	0.500	0.478	.014	0.472	.013	0.268	.010	0.459	.017
r_{g1c}	0.700	0.676	.011	0.677	.015	0.683	.017		
r_{g2c}	0.650	0.640	.011	0.655	.012	0.562	.017		
r_{p1c}	0.500	0.498	.003	0.498	.003	0.380	.003	0.500	.004
r_{p2c}	0.500	0.497	.002	0.498	.003	0.162	.002	0.213	.003

Theoretical values for genetic parameters

		1c	2c	12
all h^2 = 0.50	r_g	0.7	0.65	0.80
all σ^2_p = 100	r_p	0.50	0.50	0.50

Prediction of selection bias

If the genetic parameters in the population were known, formulae of the previous sections could be used to predict the expected selection bias for alternative methods of analysis. In general there is a considerable difference in computational requirements between AOV and ML so that use of ML is only justified when it is expected to produce markedly less biassed estimates.

Holding all other parameters constant, the theoretical influence of individual genetic parameters on the magnitude of bias can be evaluated. Figure 1 shows for a specific set of parameters the effect of the environmental correlation between c and y_2 on the expected relative bias. Unless r_{p2c} is very small so that the partial within family correlation $r_{w2c.1}$ is negative, ML is for all parameters expected to give less biassed estimates, for the sire components and h_2^2 in particular considerably so. The difference in relative bias between the two methods appears to be fairly constant over the range of intermediate values of r_{p2c}, becoming smaller with r_{p2c} tending to unity.

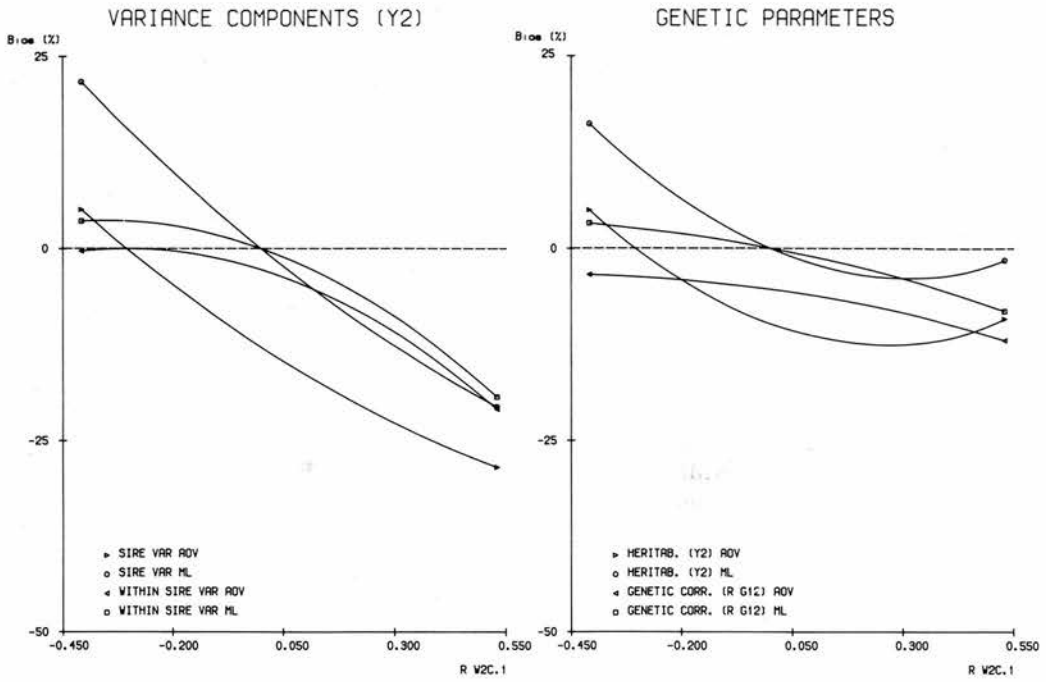
If all genetic correlations are unity, for instance, r_{g12} will be estimated unbiassedly by both ML and AOV. Bias in ML estimates of variance components and of h_2^2 will then tend to zero with increasing environmental correlation between c and y_1, for the sire components almost linearly so and independent of the level of heritabilities. In contrast, AOV estimates of variance components for y_2 and consequently h_2^2 are independent of r_{wic}, the bias decreasing with level of heritability, while the corresponding bias in σ_{w12} and σ_{s12} will increase linearly with r_{w1c}. Similar calculations were carried out for a range of constellations of genetic and environmental parameters. On the whole, ML proved to be far less affected by selection than the AOV.

In reality, however, genetic parameters for the culling variate are unkown. As shown, an analysis including survival as an all-or-none trait provides unbiassed estimates of h_c^2, h_1^2, r_{g1c}, r_{g1c} and biassed estimates of h_2^2, r_{g2c}, r_{g12} and r_{p12}. As the within family correlation between c and y_2 is not estimable, no accurate prediction of the selection bias in a practical situation and subsequent correction is possible. A crude idea of the amount of bias can be obtained, however, if an assumption about the range of values likely to cover r_{w2c} can be made. ROBERTSON (1966) for example assumed that the phenotypic and the genetic regression coefficients were equal.

Treating the estimates from the ML analysis and the assumed value of r_{w2c} as if they were the true parameters, the expected relative bias for each parameter can be obtained. Correcting each of the biassed estimates, i. e. r_{g2c}, r_{g12}, r_{p12} and h_2^2 for this bias then gives an approximation of the appropriate values before selection. This approach implicitly assumes a linear relationship between a parameter and its expected bias which is acceptable at least for small biasses. Table 7 illustrates for ML estimates that the quality of approximation is determined by the assumption about r_{w2c}. Corresponding calculations for AOV gave less good approximations even for values of r_{w2c} close to the true values as biasses were larger and the assumption of linearity of bias in the parameter value was less appropriate.

Conclusion

Even for moderate selection intensities variance/covariance estimates by AOV type procedures are likely to be severely biassed by selection. Maximum likelihood will account for selection bias when all information contributing to the selection decision is included in the model of analysis, at least for a one-way classification. Otherwise ML estimates will be biassed as well, nevertheless for a large range of genetic parameter constellations still considerably less than the corresponding AOV estimates. Thus, if data have been subject to



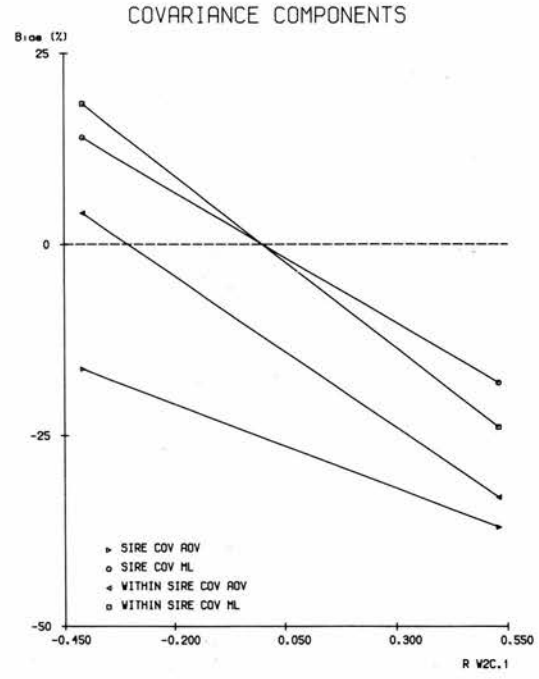


Fig. 1. Expected relative bias $((\hat{\sigma}_{ij} - \sigma_{ij}).\ 100/\sigma_{ij})$ in estimates of variance and covariance components and genetic parameters for y_2 due to selection on a correlated trait c
Culling rate 30%, all $h^2 = 0 \cdot 25$, $r_{g1c} = 0 \cdot 7$, $r_{g2c} = r_{g12} = 0 \cdot 8$, $r_{w1c} = 0 \cdot 43$, $r_{w12} = 0 \cdot 48$, $r_{w2c\cdot1} = r_{w2c} - r_{w1c}r_{w12}$

Table 7. **ML estimates of biassed genetic parameters and approximate corrections**
(Genetic parameters as in Table 2)

		h_2^2	r_{g2c}	r_{g12}	r_{p12}
True value		0.250	0.800	0.800	0.500
Expected estimate		0.224	0.743	0.788	0.458
$r_{w2c} = 0$	Bias (%)	12.8	5.5	1.2	8.8
	Corr. E.	0.199	0.704	0.774	0.421
$r_{w2c} = 0.217$	Bias (%)	0.8	0.5	0.1	0.6
$(=0.5 \times r_{w1c})$	Corr. E.	0.222	0.740	0.787	0.455
$r_{w2c} = 0.433$	Bias (%)	-7.4	-5.8	-1.6	-6.4
$(= r_{w1c})$	Corr. E.	0.242	0.789	0.800	0.489
$r_{w2c} = 0.469$	Bias (%)	-10.9	-11.1	-3.2	-10.7
(true value)	Corr. E.	0.251	0.835	0.814	0.513
$r_{w2c} = 0.640$	Bias	-11.7	-13.2	-3.9	-12.2
$(= 1.5 \times r_{w1c})$	Corr. E.	0.266	0.856	0.897	0.477

selection, an ML procedure seems to be the method of choice for analysis – provided computational requirements can be met. Further research should establish the expected selection bias for ML analysis in a mixed model situation.

Acknowledgements

This work was partly supported by the Milk Marketing Board of England and Wales, Thomas Ditton, Surrey. Numerous discussions with Dr. W. G. Hill stimulated the course of work.

Summary

The rôle of selection bias in estimating variance/covariance components by analysis of variance and maximum likelihood is investigated. Selection is considered for the general case when the selection criterion includes some unrecorded information correlated to the traits under analysis. Prediction formulae for the expected estimates are derived for a one-way classification and compared to simulation results.

Zusammenfassung

Verzerrung von Varianz- und Kovarianzschätzern als Folge einer Selektion eines korrelierten Merkmals

Der Einfluß von Selektion auf die Schätzung von Varianz- und Kovarianzkomponenten mittels einer Varianzanalyse und durch Maximum Likelihood wird untersucht. Es wird die allgemeine Situation betrachtet, in der das Selektionskriterium sich aus den zu analysierenden Merkmalen und nicht erfaßbarer, aber korrelierter Information zusammensetzt. Für eine Einwegklassifikation werden Formeln für die erwarteten Schätzwerte abgeleitet und mit Ergebnissen von simulierten Daten verglichen.

Résumé

Biais en estimateurs composants de variance et covariance du à la sélection sur un caractère corrélé

L'influence de la sélection sur l'estimation de composants de variance et covariance est examinée moyennant une analyse de variance et par le maximun likelihood. On considère la situation générale où le critère de sélection se compose de caractères à analyser et d'informations non-enregistrées mais corrélées. Les formules pour les valeurs estimatives espérées sont dérivées pour une classification one-way et comparées avec les résultats simulés.

References

AITKEN, A. C., 1934: A note on selection from a multi-variate normal population. Proc. Edinburgh Mathem. Soc. **4**, 106–110.

ANDERSON, T. W., 1973: Asymptotically efficient estimation of covariance matrices with linear structure. Ann. Stat. **1**, 135–141.

CORBEIL, R. R.; SEARLE, S. R., 1976: A comparison of variance component estimators. Biometrics **32**, 779–791.

COX, N. R., 1974: Estimation of the correlation between a continous and a discrete variable. Biometrics **30**, 171–178.

CURNOW, R. N., 1961: The estimation of repeatability and heritability from records subject to culling. Biometrics **17**, 553–556.

HARVILLE, D. A., 1977: Maximum likelihood approaches to variance component estimation and to related problems. J. A. S. A. **72**, 320–338.

HENDERSON, C. R., 1973: Sire evaluation and genetic trends. Proc. of the Animal Breeding and Genetics Symposium in honour of Dr. J. L. Lush, ASAS and ADSA, Champaign, Illinois, p. 10–41.

HENDERSON, C. R., 1975: Best linear unbiased estimation and prediction under a selection model. Biometrics **31**, 423–447.

HENDERSON, C. R.; QUAAS, R. L., 1977: A comparison of method 3 and maximum likelihood estimators of heritabilities and genetic correlations, J. Anim. Sci. **45**, Suppl. 1, 22–23.

LUSH, J. L.; SHRODE, R. R., 1950: Changes in milk production with age and milking frequency. J. Dairy Sci. **33**, 338–347.

OLAUSSON, A.; RØNNINGEN, K., 1975: Estimation of genetic parameters for treshold characters. Acta Agric. Scand. **25**, 201–208.

PATTERSON, H. D.; THOMPSON, R., 1971: Recovery of inter-block information when block sizes are unequal. Biometrika **58**, 545–554.

PEARSON, K., 1903: Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. Phil. Trans. Roy. Soc. London A **200**, 1–66.

ROBERTSON, A., 1966: A mathematical model of the culling process in dairy cattle. Anim. Proc. **8**, 95–108.

ROBERTSON, A.; LERNER, I. M., 1949: The heritability of all-or-none traits: Liability of poultry. Genetics **34**, 395–411.

ROTHSCHILD, M. F.; HENDERSON, C. R.; QUAAS, R. L., 1979: Effects of selection on variances and covariances of simulated first and second lactations. J. Dairy Sci. **62**, 996–1002.

SCHAEFFER, L. R.; SOONG, 1978: Selection bias and REML variance-covariance component estimation. J. Dairy Sci. **61**, Suppl. 1, 91–92 (Abstr.)

SEARLE, S. R., 1966: Matrix algebra for the biological sciences. New York: Wiley.

WICKHAM, B. W., 1975: Sire evaluation using selected records. Ph. D. thesis, Cornell University.

Authors' address: Inst. of Animal Genetics, Edinburgh, University, Edinburgh EH9 3JN, Scotland

24

Relationship between the cumulative different and best linear unbiased predictor methods of evaluating bulls

by

R. Thompson

Anim. Prod. 1976, **23**: 15–24

RELATIONSHIP BETWEEN THE CUMULATIVE DIFFERENCE AND BEST LINEAR UNBIASED PREDICTOR METHODS OF EVALUATING BULLS†

R. THOMPSON

Statistical Laboratory, Iowa State University, Ames, Iowa, U.S.A.
and
ARC Unit of Statistics, University of Edinburgh, 19 Buccleuch Place, Edinburgh EH8 9LN‡

SUMMARY

The relationship between two methods of sire evaluation (Bar-Anan and Sacks's cumulative difference method (CD) and Henderson's best linear unbiased predictor method (BLUP)) is shown. This relationship suggests a modification, with more theoretical justification, to the CD method.

INTRODUCTION

THE purpose of this paper is to show the relationship between two methods of sire evaluation. One is the cumulative difference method (CD) used in Israel for measuring dairy bulls' breeding values and discussed by Bar-Anan and Sacks (1974). The other method is called the best linear unbiased predictor method (BLUP) by Henderson and is discussed by him in several papers (Henderson, Kempthorne, Searle and von Krosigk (1959), and Henderson (1963 and 1973)).

DEVELOPMENT OF METHODS

In this section the alternative computational forms for evaluating bulls by the CD and BLUP methods are developed. To make this paper more readable, proofs of results that depend on matrix algebra are put in the next section.

The CD method is a method of predicting breeding values of sires when data are available only on progeny. The usual approach for evaluating sires with progeny is essentially a two-stage procedure (Robertson and Rendel, 1954; Cunningham, 1965). First unbiased estimates of the progeny means are obtained and then the breeding values of the sires are found by weighting these estimated means by their heritabilities.

Usually (Robertson and Rendel, 1954; Bar-Anan and Sacks, 1974) a model of the form

$$y_{kil} = \alpha_k + b_i + e_{kil} \tag{1}$$

is assumed, where y_{kil} is the yield of the lth daughter of sire i in herd-season k, α_k represents the effect of herd-season k, b_i is the effect of sire i and e_{kil} is a random variable with mean zero and variance σ^2. This linear model is then

† Journal Paper No. J-8267 of the Iowa Agriculture and Home Economics Experiment Station, Ames. Project No. 1669. Part of this work was done under NIH Grant 13827.

‡ Present address.

used to give estimates of the progeny means. One possibility (e.g. Cunningham, 1965) is to estimate b_i by least squares.

Once least squares estimates of the progeny means, $\hat{b}_i$, have been obtained, predicted values of the breeding values of the sires can be found by regressing these estimates of sire progeny means on the breeding values of the sires. Cunningham (1965) suggests that the breeding value of sire i be estimated by (in my notation)

$$2\hat{g}_i = \left(\frac{\sigma_A^2}{2}\right)\left[\left(\frac{\sigma_A^2}{4}\right)+\sigma_{e_i}^2\right]^{-1}\hat{b}_i \tag{2}$$

where σ_A^2 is the additive genetic variance, and $\sigma_{e_i}^2$ is the sampling variance of $\hat{b}_i$. We use $2\hat{g}_i$ to represent the predicted breeding value of sire i, so that $\hat{g}_i$ represents the predicted value of a daughter of sire i. The variance σ^2 is usually written in terms of σ_P^2, the phenotypic variance, and σ_A^2. From (1), $2b_i$ can be thought of as the breeding value of sire i, and taking this to be a random variable with variance σ_A^2, the phenotypic variance can be written as

$$\sigma_P^2 = \sigma^2+(\tfrac{1}{4})\sigma_A^2. \tag{3}$$

When there are no herd-season effects, and there are n progeny records on sire i, (2) reduces using (3) and writing h^2 for σ_A^2/σ_P^2, the heritability, to

$$2\hat{g}_i = [2nh^2/(4+(n-1)h^2)]\hat{b}_i \tag{4}$$

i.e. the predicted breeding value is found by regressing the estimated progeny value using a coefficient $2nh^2/(4+(n-1)h^2)$.

The least squares equations are now written in an alternative form using, when possible, the notation of Bar-Anan and Sacks (1974). The estimates of sire effects, $\hat{b}_i$, satisfy (from (22), next section)

$$w_i\hat{b}_i = w_iC_i+w_iA_{i'}, \tag{5}$$

where w_i is the weight or effective number of daughters (Robertson and Rendel, 1954) associated with sire i. w_i is the sum of weights, w_{ik}, from the individual herd-seasons and

$$w_{ik} = n_{ik}(1-n_{ik}/n_k) \tag{6}$$

where n_{ik} is the number of daughters of sire i in herd-season k and n_k is the total number of daughters in herd-season k. The average yield of the n_{ik} daughters of sire i in herd-season k is $\bar{D}_{ik}$ and

$$w_iC_i = \sum_k w_{ik}(\bar{D}_{ik}-\bar{D}_{i'k}) \tag{7}$$

where $\bar{D}_{i'k}$ is the average yield of the contemporary herd mates of the daughters of sire i in herd-season k, and $\bar{D}_{ik}-\bar{D}_{i'k}$ is the difference between the mean of the daughters of sire i and the mean of the contemporaries in herd-season k.

The term $A_{i'}$ is similar to term (9) in Bar-Anan and Sacks (1974) and is given by

$$w_iA_{i'} = \sum_k w_{ik}\Big(\sum_{j\neq i} n_{jk}\hat{b}_j\Big)/(n_k-n_{ik}) \tag{8}$$

and is a measure of the genetic merit of the contemporaries of sire i. Since $A_{i'}$ is a function of the estimates $\hat{b}_i$, the least squares equations can be difficult to solve especially if there are many bulls. Robertson and Rendel

(1954) suggest making an approximation to the least squares solution by ignoring the last term in (5) and estimating b_i by

$$b_i^\star = C_i. \tag{9}$$

The estimates of breeding values, from these estimates of b_i, are usually found (Bar-Anan and Sacks, 1974) by replacing $\hat{b}_i$ and n in (4) by $b_i^\star$ and w_i.

As a consequence the predicted performances of future daughters, $g_i^\star$, satisfy

$$\begin{aligned} g_i^\star &= [w_i h^2/(4+(w_i-1)h^2)]b_i^\star \\ &= [w_i/(w_i+\gamma)]b_i^\star \end{aligned} \tag{10}$$

where $\gamma = (4-h^2)/h^2$.

From (10), the estimates $b_i^\star$ are regressed back by a factor

$$w_i h^2/(4+(w_i-1)h^2)$$

to give the $g_i^\star$. Because $b_i^\star$ and $g_i^\star$ are based on C_i and this is a weighted sum of deviations of a bull's daughters from their contemporaries, the method is called the contemporary comparison (CC) method. Both $b_i^\star$ and $g_i^\star$ have been called contemporary comparisons in the literature but I shall try to differentiate $b_i^\star$ from $g_i^\star$.

When the CC method was initially suggested (1954), it was thought reasonable to neglect the term $A_{i'}$ in (5), but it has been recognized that this has disadvantages. For instance, as a bull grows older the contemporaries of his daughters are likely to be daughters of progressively younger bulls. Then, if selection of bulls is effective, the genetic merit of the contemporaries of the bull's daughters would be expected to increase, and the bull's contemporary comparison ($b_i^\star$ or $g_i^\star$) to decrease as the bull grows older. This, in fact, happens and Bar-Anan and Sacks (1974) give an example. There will also be difficulties if contemporary comparisons ($b_i^\star$ or $g_i^\star$) are used to measure genetic trends over time.

Bar-Anan and Sacks (1974) describe a method that tries to take account of this deficiency in the CC method. Their estimate is called the cumulative difference (CD) and is made up of two parts, the contemporary comparison ($g_i^\star$) and an adjustment for the average genetic merit of the contemporaries of the bull's daughters.

Since information on a bull's progeny becomes available over a period of time, Bar-Anan and Sacks estimate a bull's worth at different times, t, using the available information up to this time; this they call the cumulative difference (CD_t) at time t. The predicted breeding values given by Bar-Anan and Sacks are of the form

$$CD_{it} = g_{it}^\star + A_{i't}^\star \tag{11}$$

where g_{it} is the contemporary comparison of sire i at time t and $A_{i't}^\star$ is an estimate of a weighted average of breeding values of the contemporaries of the daughters of sire i, using the estimates of breeding value available at time t, $CD_{i(t-1)}$, and is given by

$$w_i A_{i't}^\star = \sum_k w_{ik}\Big(\sum_{j \neq i} n_{jk} CD_{j(t-1)}\Big)/(n_k - n_{ik}). \tag{12}$$

The term $A^{\star}_{i't}$ is similar to the term from predicted herdmates neglected in approximating $\hat{b}_i$ by $b^{\star}_i$ but Bar-Anan and Sacks use $A^{\star}_{i't}$ as an adjustment to $g^{\star}_i$, the regressed value of $b^{\star}_i$, and I see no justification for doing this.

Dempfle (1976) has noted this difficulty in the CD method and suggests a modification. He effectively uses (5) to estimate b_i at time t, his estimate being denoted by CA_i. He uses as an estimate of $w_iA_{i'}$ a term similar to (12) with $CD_{j(t-1)}$ replaced by the estimate of b_j available at that time ($CA_{j(t-1)}$). He then regresses his estimate of b_i back, as in (10), to give a modified cumulative difference. This modification is now used by Bar-Anan (1976).

In order to investigate the CD method further, it is now compared with the BLUP method of Henderson. This method arose out of a consideration of the mixed model similar to (1) when the sire effects are thought of as random effects with variance $(\sigma^2_A/4)$. The observations are then not independent, there is a covariance of $(\sigma^2_A/4)$ between daughters of the same bull (half-sibs) and so the best linear unbiased estimates of the herd-season effects, $\hat{\alpha}_i$, satisfy weighted least squares equations. These equations can be difficult to set up, requiring the inversion of a $n \times n$ matrix representing the variances and covariances of the observations. Henderson in Henderson *et al.* (1959) has shown that the equation that α_i satisfy can be written in a form similar to the least squares equations when the sire effects are assumed to be fixed effects. He further noted that a term β_i, that arose naturally in these modified least squares equations, can be interpreted as a predictor of the breeding value of sire i. Later, Henderson (1963) showed that this term β_i is the best linear unbiased predictor of b_i and so this method has been called the BLUP method (Henderson, 1973).

The equations for β_i can be written in a form similar to that in which the least squares equations were written (equation 5). β_i then satisfies (from (27), next section)

$$(w_i+\gamma)\beta_i = w_iC_i+w_iA_{i'} \tag{13}$$

where $A_{i'}$ is given by (8) with β_j replacing $\hat{b}_j$.

It should be noted that if the last term in (13) is ignored, $\beta_i = g^{\star}_i$, (from (10)), so that the CC method could be thought of as an approximation of the BLUP method.

Equation (13) suggests a modified cumulative difference, $CD^{\star}_t$, similar to that of Bar-Anan and Sacks, where $CD^{\star}_{it}$ is an approximation to β_i and satisfies

$$(w_i+\gamma)CD^{\star}_{it} = w_iC_i+w_i\,A^{\star}_{i't} \tag{14}$$

with $A^{\star}_{i't}$ evaluated using (12) and replacing $CD_{i(t-1)}$ by $CD^{\star}_{i(t-1)}$.

DERIVATION OF RESULTS

In this section some of the results quoted in the previous section without proof are derived. The derivations depend heavily on matrix algebra. Any reader willing to accept without proof the results of the first section can go straight to the next section where the results given in the first section are discussed.

If there are $n(=\sum_{i,k} n_{ik})$ records, r herd-season effects and s sires, the

model (1) can be written in a convenient matrix form

$$\boldsymbol{y} = \boldsymbol{X\alpha} + \boldsymbol{Zb} + \boldsymbol{e} \tag{15}$$

where $\boldsymbol{y}$, $\boldsymbol{\alpha}$ and $\boldsymbol{b}$ are $n \times 1$, $t \times 1$ and $s \times 1$ vectors representing the records, herd-season effects and sire effects. Here, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are $n \times t$ and $n \times s$ design matrices with elements X_{mk} equal to one if the mth record is in herd-season $k (m = 1, \ldots, n;\ k = 1, \ldots, r)$ and zero otherwise, and elements Z_{mi} equal to one if the mth individual is a progeny of sire $i (m = 1, \ldots, n;\ i = 1, \ldots, s)$ and zero otherwise.

The least squares estimates $\boldsymbol{\alpha}$ and $\boldsymbol{b}$ satisfy

$$\begin{aligned} \boldsymbol{X'X\hat{\alpha}} + \boldsymbol{X'Z\hat{b}} &= \boldsymbol{X'y} \\ \boldsymbol{Z'X\hat{\alpha}} + \boldsymbol{Z'Z\hat{b}} &= \boldsymbol{Z'y} \end{aligned} \tag{16}$$

and eliminating $\hat{\boldsymbol{\alpha}}$ from (16) we get

$$\boldsymbol{Z'SZ\hat{b}} = \boldsymbol{Z'Sy}, \tag{17}$$

where $\boldsymbol{S} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'}$ and $\boldsymbol{I}$ is the identity matrix.

It is probably useful to interpret the terms of (16) and (17) algebraically. The matrix $\boldsymbol{X'X}$ is a diagonal matrix because a record is in only one herd-season and has elements n_k; similarly $\boldsymbol{Z'Z}$ is a diagonal matrix with elements $\sum_k n_{ik}$. Now, $\boldsymbol{X'Y}$ and $\boldsymbol{Z'Y}$ represent the totals of records over herd-seasons and sires respectively. For instance, $(\boldsymbol{Z'y})_i = \sum_k n_{ik}\bar{D}_{ik}$. The diagonal elements of $\boldsymbol{Z'SZ}$, are

$$\sum_k (n_{ik} - n_{ik}^2/n_k) = \sum_k w_{ik} = w_i,$$

the effective number of daughters of sire i (equation 6). The non-diagonal elements of $\boldsymbol{Z'SZ}$, $(\boldsymbol{Z'SZ})_{ij} (i \neq j)$ are

$$-\sum_k (n_{ik}n_{jk})/n_k,$$

and

$$(\boldsymbol{Z'Sy})_i = \sum_k (n_{ik}\bar{D}_{ik} - n_{ik}(\sum_j n_{jk}\bar{D}_{jk})/n_k). \tag{18}$$

The term $\sum_j n_{jk}\bar{D}_{jk}$ is the total yield of all animals in herd-season k. If use is made of the fact that the average yield of contemporaries of i, $D_{i'k}$, satisfies $(n_k - n_{ik})\bar{D}_{i'k} = \sum_{j \neq i} n_{jk}\bar{D}_{jk}$, (18) can be written as

$$\begin{aligned} (\boldsymbol{Z'Sy})_i &= \sum_k (w_{ik}\bar{D}_{ik} - n_{ik}(n_k - n_{ik})\bar{D}_{i'k}/n_k) \\ &= \sum_k w_{ik}(\bar{D}_{ik} - \bar{D}_{i'k}). \end{aligned} \tag{19}$$

This is equal to the effective number of daughters times the contemporary comparison $b_i^\star$ (equations (9) and (7)).

The variance, $\sigma^2_{e_i}$, of the least squares estimate is $((\boldsymbol{Z'SZ})^{-1})_{ii}$, the ith

diagonal element of $\boldsymbol{Z'SZ}^{-1}$, so that the breeding value $\hat{g}_i$, found by regressing $\hat{b}_i$ according to (2), is

$$2\hat{g}_i = \left(\frac{\sigma_A^2}{2}\right)\left[\left(\frac{\sigma_A^2}{4}\right) + ((\boldsymbol{Z'SZ})^{-1})_{ii}\sigma^2\right]^{-1} \hat{b}_i \tag{20}$$

with the vector of breeding values of all the sires estimated by

$$2\hat{\boldsymbol{g}} = \left(\frac{\sigma_A^2}{2}\right)\left[\left(\frac{\sigma_A^2}{4}\right)\boldsymbol{I} + ((\boldsymbol{Z'SZ})^{-1})^{\delta}\sigma^2\right]^{-1} \hat{\boldsymbol{b}}, \tag{21}$$

where $((\boldsymbol{Z'SZ})^{-1})^{\delta}$ is a diagonal matrix with ith diagonal element $(\boldsymbol{Z'SZ})^{-1})_{ii}$.

Equation (17) can be written in the alternative form

$$\boldsymbol{W}\hat{\boldsymbol{b}} = \boldsymbol{Z'Sy} + (\boldsymbol{W} - \boldsymbol{Z'SZ})\hat{\boldsymbol{b}} \tag{22}$$

where $\boldsymbol{W}$ is a diagonal matrix with elements the diagonal elements of $\boldsymbol{Z'SZ}$. These diagonal elements have been shown to be w_i, the number of effective daughters of sire i. Since (17) can be difficult to solve, one possibility is to ignore the last term in (22) and estimate $\hat{\boldsymbol{b}}$ by

$$\boldsymbol{b}^{\star} = \boldsymbol{W}^{-1}\boldsymbol{Z'Sy}. \tag{23}$$

Using (19) and (7), this is a matrix form of Robertson and Rendel's estimate (1954), (equation (9) above).

The term ignored in (22), $\boldsymbol{W}^{-1}(\boldsymbol{W} - \boldsymbol{Z'SZ})\hat{\boldsymbol{b}}$ is a vector and the element related to sire i can be written as

$$\begin{aligned}(\boldsymbol{W}^{-1}(\boldsymbol{W} - \boldsymbol{Z'SZ})\hat{\boldsymbol{b}})_i &= (\sum_k (n_{ik}(\sum_{j \neq i} n_{jk}\hat{b}_j)/n_k))/(\sum w_{ik}) \\ &= (\sum_k w_{ik}(\sum_{j \neq i} n_{jk}\hat{b}_j)/(n_k - n_{ik}))/(\sum_k w_{ik}).\end{aligned} \tag{24}$$

Using (22), (24) and (8) we now see that (5) is an alternative form of the least squares equation (17).

The results related to the BLUP procedure can now be derived. In this procedure the observations are supposed to have a variance matrix $\boldsymbol{V}$ given by

$$\boldsymbol{V} = \boldsymbol{I}\sigma^2 + \boldsymbol{ZZ'}(\sigma_A^2/4)$$

indicating that the covariance between daughters of the same bull (half-sibs) is $\sigma_A^2/4$, as expected. One way of generating such a variance matrix is to think of the sire effects in (15) as random variables with variance $(\sigma_A^2/4)$. With this model Henderson in Henderson *et al.* (1959), has shown that the best linear unbiased estimate of $\boldsymbol{\alpha}$ satisfies

$$\begin{aligned}\boldsymbol{X'X\alpha} + \boldsymbol{X'Z\beta} &= \boldsymbol{X'y} \\ \boldsymbol{Z'X\alpha} + (\boldsymbol{Z'Z} + \boldsymbol{I}\gamma)\boldsymbol{\beta} &= \boldsymbol{Z'y}\end{aligned} \tag{25}$$

which can be more convenient than the usual form

$$\boldsymbol{\alpha} = (\boldsymbol{X'V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X'V}^{-1}\boldsymbol{y},$$

since it avoids inverting the $n \times n$ matrix $\boldsymbol{V}$.

Henderson has emphasized (1963, 1973) that $\boldsymbol{\beta}$ are not just computational artefacts but can be interpreted as the best linear unbiased predictor (BLUP) of $\boldsymbol{b}$ (under this mixed model).

Eliminating $\boldsymbol{\alpha}$ from (25), gives

$$(\boldsymbol{Z'SZ}+\boldsymbol{I}\gamma)\boldsymbol{\beta} = \boldsymbol{Z'Sy}. \tag{26}$$

In general the predicted values from (26), $\boldsymbol{\beta}$, will differ from those from (21), $\hat{\boldsymbol{g}}$. As one might expect, however, if there are no fixed effects, $\boldsymbol{\beta}$ reduces to $\hat{\boldsymbol{g}}$. One way of differentiating the two approaches is by comparing (16) with (25) from the point of view of estimating $\boldsymbol{\alpha}$; (16) is consistent with estimating $\boldsymbol{\alpha}$ with model (15) with $\boldsymbol{b}$ as fixed sire effects, while (25) is consistent with model (15) with $\boldsymbol{b}$ as random variables with variance $\sigma_A^2/4$. If $h^2 = 4$, then equations (25) are equivalent to (16), so in one sense (16) takes too much account of the genetic structure since by definition h^2, heritability, is less than or equal to one.

The rewriting of (17) in the form of equation (22) suggests an alternative form for (26) i.e.

$$(\boldsymbol{W}+\boldsymbol{I}\gamma)\boldsymbol{\beta} = \boldsymbol{Z'Sy}+(\boldsymbol{W}-\boldsymbol{Z'SZ})\boldsymbol{\beta}. \tag{27}$$

Again using (24) to relate A_i, to $(\boldsymbol{Z'SZ}-\boldsymbol{W})\boldsymbol{\beta}$ we find that (27) is the matrix analogue of (13).

DISCUSSION

Relationship of CD to BLUP

Rewriting the scheme suggested by the BLUP procedure (14) as

$$\begin{aligned} CD_{it}^{\star} &= g_{it}^{\star}+(w_i/(w_i+\gamma))A_{i't}^{\star} \\ &= (w_i/(w_i+\gamma))(b_{it}^{\star}+A_{i't}^{\star}) \end{aligned}$$

shows that this differs from the CD scheme (11) in that both the contemporary comparison, $b_{it}^{\star}$, and the measure of contemporary performance, $A_{it}^{\star}$, are regressed back, while in Bar-Anan and Sacks's scheme only $b_{it}^{\star}$ is regressed back. The regression-back effect is more important when there are fewer daughters; for example, with 100 and 500 effective daughters and a heritability of 0·2, the regression coefficients are of the order of 0·8 and 0·95 respectively.

The CD method cannot be justified by recourse to the BLUP method. Neither can it be justified, as noted in the first section, in terms of the two-stage procedure of estimating fixed effects and regressing these back. There seems, therefore, little theoretical basis for the method as stated by Bar-Anan and Sacks. Simple modifications, such as (14) and the one suggested by Dempfle (1976), have a much stronger backing.

For the modified scheme (14) a fairly obvious suggestion is, once an estimate $CD_{it}^{\star}$ is found from (14), to repeat the process substituting $CD_{it}^{\star}$ for $CD_{i(t-1)}^{\star}$ to give another estimate of $CD_{it}^{\star}$. This might be more useful early in a bull's life when $CD_{it}^{\star}$ should be an improvement on $CD_{i(t-1)}^{\star}$.

If the scheme is iterated until the values stabilize the resulting estimates are BLUP estimates. The number of iterations needed until convergence obviously depends on the structure and number of observations and so it is difficult to give any definite rule for the number of such iterations.

Dempfle (1976) suggests that his scheme might be iterated and this leads to least squares estimates of sire effects b_i which are then regressed back to

give estimates of breeding values. Dempfle's modification is similar to (14) in that both the contemporary comparison and the adjustment are regressed back. It differs from (14) in that in (14) the adjustment is using *regressed* estimates whilst in Dempfle's scheme *unregressed* estimates are used. Because of the relationship of Dempfle's scheme to the two-stage procedure and (14) to the BLUP procedure and the fact that BLUP is better than the two-stage procedure I prefer (14) to Dempfle's modification.

Constraints

It is noted that (5) does not define the least squares estimates $\hat{b}_i$ uniquely, for if $\hat{b}_i$ satisfies (5) and a constant is added to each element these still satisfy (5).

From a practical point of view this does not matter since we are usually interested in differences between sires. $\hat{b}_i$ is found to satisfy simple constraints, either that the sum of b_i is zero, or that a weighted sum of b_i is zero, or that one particular value of b_i is given an assigned value (e.g. Searle (1971), section 5.7).

There can be difficulty in comparing these least squares estimates $\hat{b}_i$ at different ages of the bulls if the constraints on $\hat{b}_i$ are not chosen appropriately. For instance, if $\hat{b}_{it}$ is a solution to (4) using all data up to time t with the very common constraint

$$\sum_{i=1}^{s} \hat{b}_{it} = 0,$$

and if s' sires are evaluated at time $t+1$ with the similar constraint

$$\sum_{i=1}^{s'} \hat{b}_{i(t+1)} = 0,$$

then

$$\sum_{i=1}^{s} \hat{b}_{i(t+1)} = -\sum_{i=s+1}^{s'} \hat{b}_{i(t+1)}.$$

If the bulls $(s+1)$ to (s') are better than bulls 1 to s, as might be expected since they will be younger bulls, then the estimated values of progeny means for bulls 1 to s will decrease from time t to $(t+1)$. This suggests that a more appropriate procedure is to constrain the estimate of one particular sire (or the sum of estimates of a group of sires) to one value for all time. Then the estimates $\hat{b}_{it}$ are effectively estimates of the difference between sire i and the reference sire(s) at time t. A similar argument could be used to explain the decrease in the contemporary comparisons ($b_i^\star$ and $g_i^\star$) of bulls as they grow older, since the sum of $w_i C_i$, the deviations of bull's daughters from their contemporaries, is zero and so

$$\sum w_{it} b_{it}^\star = 0,$$

where $b_{it}^\star$ is $b_i^\star$ evaluated using data available at time t and w_{it} is the effective number of daughters at time t.

In the discussion of the BLUP method a mixed model of the form (1) was used primarily to highlight the relationship with the CD method. There are difficulties with this model, however, since from (25) we find that

$$\Sigma \beta_i = 0,$$

so that again, as bulls age, their predicted values decrease. This is really a fault of the model, not of the method. Henderson (1973) suggests adding some group fixed effects to the model where the group is defined to be a set of sires entering service at the same time. Equations similar to (14) could be developed for this model. Alternatively, if bulls are sons of tested bulls the variance matrix could be changed to take account of the relationships.

Updating records

In the present paper, the cumulative difference has been interpreted as including *all* the information up to time t and this seems the best interpretation of CD_{it}. This is how the method is implemented in Israel (R. Bar-Anan, personal communication). Bar-Anan and Sacks seem to imply (1974, p. 62), however, that the cumulative difference should be calculated for each time period using just information in that time period and that the estimate for the breeding value of a bull is found by averaging CD_{it} across time. This will have the undesirable effect of regressing back the parts of $b_{it}^{\star}$ from each period more than if an estimate of $b_{it}^{\star}$ is found from all the time periods and then regressing it back to give $g_{it}^{\star}$. For example, ignoring the correction term A_{it}, if the estimate $b_{it}^{\star}$ of progeny mean for sire i using all the data up to time t is equal to b for periods $t = 1, \ldots, T$, and the effective number of daughters in each period is w then

$$b_{it}^{\star} = b$$

and

$$g_{it}^{\star} = wTb/(wT+\gamma),$$

whereas the estimate found by averaging the estimates of breeding value from each period separately is $wb/(w+\gamma)$. This can be considerably less than $g_{it}^{\star}$. For instance, if $w = 50$, $T = 10$ and heritability is 0·2, the ratio of the alternative estimates is approximately three-quarters. Dempfle (1976) makes the same point. This is related to the fact that although a daughter's record appears in only one time period, sire tests in different periods are not independent, since there is a covariance between the records of a bull's daughters in different periods.

It might be more efficient to update sequentially the terms used in evaluating the bulls rather than work out all the quantities afresh whenever bulls are evaluated.

ACKNOWLEDGEMENT

I am grateful for the comments of Professor O. Kempthorne, expecially for pointing out the difficulty in interpreting $\hat{b}$.

REFERENCES

Bar-Anan, R. 1976. The modified cumulative difference. *Proc. 27th Mtg EAAP, Zurich, Switzerland.*

Bar-Anan, R. and Sacks, J. M. 1974. Sire evaluation and estimation of genetic gain in Israeli dairy herds. *Anim. Prod.* **18**: 59–66.

Cunningham, E. P. 1965. The evaluation of sires from progeny test data. *Anim. Prod.* **7**: 221–231.

Dempfle, L. 1976. A note on the properties of the cumulative difference method for sire evaluation. *Anim. Prod.* **23**: 121–124.

HENDERSON, C. R. 1963. Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding* (ed. W. D. Hanson and H. F. Robinson), pp. 141–163. Publ. 982. National Academy of Sciences, National Research Council, Washington, D.C.

HENDERSON, C. R. 1973. Sire evaluation and genetic trends. In *Proc. Anim. Breed. Genet. Symp., Blacksburg, Virginia*, pp. 10–41. American Society of Animal Science, Champaign, Illinois.

HENDERSON, C. R., KEMPTHORNE, O., SEARLE, S. R. and VON KROSIGK, C. M. 1959. The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, **15**: 192–218.

ROBERTSON, A. and RENDEL, J. M. 1954. The performance of heifers got by artificial insemination. *J. agric. Sci., Camb.* **44**: 184–192.

SEARLE, S. R. 1971. *Linear Models.* Wiley, London.

(*Received* 12 *August* 1975)

25

Sire Evaluation

by

R. Thompson

BIOMETRICS 35, 339–353
March, 1979

Sire Evaluation

R. THOMPSON

ARC Unit of Statistics, University of Edinburgh, Mayfield Road, Edinburgh, EH9 3JZ, Scotland

Summary

Recent developments in sire evaluation procedures are reviewed and relationships of several methods of sire evaluation are indicated. The Best Linear Unbiased Predictor (BLUP) approach is discussed. This gives a convenient framework for discussing recent advances in variance component estimation. Other topics discussed include selection and the use of relationships between animals. Areas needing further work are indicated; these include the estimation of genetic trend and evaluation methods for non-linear models.

1. Introduction

This paper reviews developments in sire evaluation in the last ten years. Most of these developments have arisen in dairy cattle and we use this as an example although, in principle at least, the methods could be used in other species.

In Section 2 we discuss several contemporary comparison methods. One important development has been the introduction of the Best Linear Unbiased Predictor (BLUP) approach by Henderson. We discuss this method in Section 3, noting the relationship with selection index theory and with least squares approaches. In Section 4 we discuss some of the implications of selection and in particular a selection model of Henderson (1975). We consider how relationships between sires are used in evaluating sires in Section 5. In Section 6 we consider the estimation of genetic trends and suggest the relationship between the BLUP approach and methods used in selection experiments are worth investigating.

Most sire evaluation methods depend on using estimates of variance components and so we discuss in Section 7 recent developments in this area. We find that the BLUP approach gives a convenient way of describing these advances. Most methods of sire evaluation are based on linear models but sometimes non-linear models are more appropriate. We note in Section 8 that methods for fixed effect non-linear models might be extended to give methods of sire evaluation for non-linear models.

2. Methods of Sire Evaluation

In this section we compare several methods used for sire evaluation. We show how the methods have evolved and illustrate the common features of the methods. Up to 25 years ago most methods of evaluating bulls used dam-daughter comparisons. There was then an introduction of methods based on comparing daughters with their herdmates. This introduction coincided with the spread of artificial insemination, which gave the possibility of having daughters of a bull in many herds.

Key words: Sire evaluation; Best Linear Unbiased Prediction; Selection; Relationship matrix; Variance components.

We first use as an illustration the contemporary comparison scheme used until recently in England and Wales. This scheme uses only first lactations and assumes a model of the form

$$y_{ijk} = \beta_i + u_j + e_{ijk} \tag{1}$$

where y_{ijk} represents the heifer yield of the k*th* daughter of the j*th* sire in the i*th* herd-year-season. There are n_{ij} daughters of the j*th* sire in the i*th* herd-year-season. We assume that e_{ijk} is a random variable with mean zero and variance σ^2. β_i represents the effect of the i*th* herd-year-season and u_j the effect of the j*th* sire. The sire effects have been interpreted in two ways. We might assume u_j are fixed effects and then (1) represents a fixed effect model. Alternatively we might assume that the u_j are independently distributed with variance σ^2_b. The effects u_i are now random effects and (1) represents a mixed model. The phenotypic variance, σ^2_p, equals $\sigma^2 + \sigma^2_b$ and the covariance between daughters of the same bull, σ^2_b, is assumed to be $\sigma^2_A/4$, where σ^2_A is the additive genetic variance. In this section we will use the fixed effect model and the variance parameters from the mixed model. The use of the mixed model will be discussed in Section 3.

One way of predicting the breeding value of sires is by a two-stage procedure of first estimating the progeny means and then regressing these estimates of progeny means on the breeding values of the sires (Searle 1964*a*, Cunningham 1965). For instance, Cunningham (1965) suggests that the predicted value of a daughter of sire j be given by

$$\hat{g}_j = \sigma^2_b[\sigma^2_b + \sigma^2_{ej}]^{-1}\,\hat{u}_j \tag{2}$$

where $\hat{u}_j$ is a least squares estimate found using the fixed effect version of model (1) and σ^2_{ej} is the sampling variance of the least squares estimate $\hat{u}_j$. The least squares estimates $\hat{u}_j$ represent sire effects corrected for herd-year-season effects and satisfy equations of the form

$$n_{i0}\beta_i + \sum_j n_{ij}\hat{u}_j = y_{i00}\,, \tag{3}$$

$$\sum_i n_{ij}\beta_i + n_{0j}\hat{u}_j = y_{0j0}\,, \tag{4}$$

where 0 indicates summation over a suffix and so n_{0j} and y_{0j0} represent the total number of daughters of the j*th* sire and their total yield. When there are no herd-year-season effects and n progeny records $\hat{g}_j$ can be written as

$$\hat{g}_j = [j/(n + \gamma)]\hat{u}_j, \tag{5}$$

where $\gamma = (4 - h^2)/h^2$ and h^2, the heritability, $= \sigma^2_A/\sigma^2_p$. We can substitute for $\hat{\beta}_i$ in (4) using (3) to give equations in $\hat{u}_j$ alone, i.e.,

$$\left[n_{0j} - \sum_i (n_{ij}^2/n_{i0})\right]\hat{u}_j - \sum_{k \neq j}\left[\sum_i (n_{ij}n_{ik}/n_{i0})\hat{u}_k\right] = y_{0j0} - \sum_i (n_{ij}y_{i00}/n_{i0})\,. \tag{6}$$

This can be written (Thompson 1976) as

$$w_j\hat{u}_j = \sum_i w_{ij}\,(\bar{D}_{ij} - \bar{D}_{i'j}) + a_j \tag{7}$$

where $a_j = \sum_{k \neq j}[\sum_i (n_{ij}n_{ik}/n_{i0})]\hat{u}_k$ can be thought of as a measure of the genetic merit of the contemporaries of the daughters of sire j and w_j is the sum of weights, w_{ij}, from the individual herd-year-seasons and $w_{ij} = n_{ij}(1 - n_{ij}/n_{i0})$. $\bar{D}_{ij}$ and $\bar{D}_{i'j}$ are the average yields of the daughters of sire j and of their contemporaries in herd-year-season i. The weight w_{ij}, sometimes called the effective number of daughters (Robertson and Rendel 1954), is inversely proportional to the variance of $(\bar{D}_{ij} - \bar{D}_{i'j})$.

Equation (6) will be difficult to solve especially if there are many bulls. Robertson and Rendel (1954) suggested making an approximation to the least squares solution by ignoring the last term in (7) and approximating u_j by C_j so that

$$w_j C_j = \sum_i w_{ij}(\bar{D}_{ij} - \bar{D}_{i'j}) . \tag{8}$$

Since C_j is based on a weighted deviation of a bull's daughters from their contemporaries, this method has been called the contemporary comparison method (CC). Predicted values of future daughters, g_i^*, are usually found (Bar-Anan and Sacks 1974) by replacing $\hat{u}_j$ and n by C_j and w_j in (5) to give $g_j^* = [w_j/(w_j + \gamma)]C_j$.

In the next section we discuss the Best Linear Unbiased Prediction (BLUP) approach of Henderson. It can be shown (compare equations (3), (4), (7) and (9) with equations (13) and (14)) that the BLUP predictions for model (1) satisfy

$$(w_j + \gamma)\hat{u}_j = \sum_i w_{ij}(\bar{D}_{ij} - \bar{D}_{i'j}) + a_j \tag{9}$$

and if we again ignore the last term in (9), representing the genetic merit of contemporaries, we see that g_i^* can be thought of as an approximate BLUP solution.

It was thought reasonable to omit the adjustment, a_j, for contemporaries genetic merit in the solution of (7) when the method was initially suggested. It was acceptable at first but later led to difficulties. For example, as a bull ages, the contemporaries of his daughters are likely to be the daughters of progressively younger bulls; hence, if selection of young bulls is effective the genetic merit of the contemporaries should increase and the bull's contemporary comparison C_j or g_j^* should decrease. Bar-Anan and Sacks (1974) give an example of this. Further genetic and environmental trends over time derived from these contemporary comparisons will be difficult to interpret.

Bar-Anan and Sacks (1974) point out that usually sires will be evaluated at discrete intervals and suggest that instead of ignoring a_j in (7) one should evaluate it using estimates of u_j available at that time. Bar-Anan (1976) calls this a Modified Cumulative Difference (MCD) procedure. Obviously one can iterate this procedure (or an equivalent BLUP scheme) until one finds a solution to (6) or (8).

Searle (1964*a*) compares the CC method with methods used in New York State and New Zealand. These schemes are based on regressing daughter performance on contemporary performance. Herdmate comparisons of the same form as (7) are used with a_j, now of the form $b\sum_i w_{ij}(\bar{D}_{i'j} - \mu)$, where b is a regression coefficient of the order of 0.1 and μ is the breed mean. These comparisons are regressed using (5) to give predicted values. Again it was found that trends in genetic merit of contemporaries made these methods invalid.

Another difficulty with the model is that we have assumed all bulls are distributed about a mean of zero and if selection of bulls is effective then younger bulls should be of higher merit. One possibility is to regress the least square estimates to a group mean, rather than to zero. The group could be defined to be a set of sires from one stud entering service at one time. Another approach takes account of the fact that if bulls are sons of tested bulls then bull effects are correlated with one another. We will see (Section 3) that the BLUP method can be easily changed to take account of these relationships.

Another scheme in common use is that of the United States Department of Agriculture (USDA) (Dickinson, Norman, Powell, Waite and McDaniel 1976). They use all lactations, corrected for age and length, and assume that all lactations have equal phenotypic and genetic variances, that environmental covariances between lactations are equal and also that genetic correlations between lactations are one. They also assume a random herd-sire

interaction. Henderson (1973) has set up a BLUP computing strategy for this model but he admits this is only feasible for small numbers of animals and USDA instead use a scheme similar to the MCD. For each record two contemporary averages are formed from heifer lactations and from later lactations. Weighted averages of these give a contemporary average for each lactation. For a contemporary average of heifer lactations, most weight is given to heifer contemporaries and only a nominal weight to cow contemporaries, and vice versa for contemporary averages of cow lactations. A weighted average of the deviations of a bull's daughters from their contemporaries is formed first combining a particular cow's deviations, then combining cows within herds and then combining over all herds. As in the MCD an adjustment for the genetic merit of the contemporaries is made using previous estimates of the contemporary comparison. Again the procedure is repeated until the contemporary comparisons stabilize. The bulls are then grouped according to their predicted value based on their sire and maternal grandsire information, group means calculated and predicted values for bulls found by regressing the contemporary comparisons back to the group means. The regression coefficient is based on (2), using an approximation for the variance of the contemporary comparison based on the number and length of lactations of both the bulls' daughters and their contemporaries and also on the number and average repeatability of the contemporaries' sires. This is slightly different from the MCD scheme where the regression coefficient does not depend on the number or average repeatability of the contemporaries' sires. Whether this difference is practically important I do not know.

The regressed least squares approach has provided a convenient framework to illustrate the relationships between several sire evaluation methods. However, this method has several drawbacks as Henderson (1978) recently emphasized. One difficulty is that the covariance structure is ignored when the sire effects and the herd-year-season effects are estimated. In the next section we discuss a method that does use the covariance structure and has more theoretical justification.

3. Best Linear Unbiased Prediction

We will use matrix notation in order to express the results compactly and avoid a plethora of suffixes and summation symbols.

For ease of exposition we will use a simple two-factor mixed model of the form

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{10}$$

where $\mathbf{y}$ is a $(n \times 1)$ vector of observations on one trait, $\mathbf{X}$ and $\mathbf{Z}$ are $n \times t$ and $n \times b$ matrices and are assumed to be of full rank. β and $\mathbf{u}$ are vectors of size t and b representing the unknown fixed and random effects. We assume $E(u_i) = 0$ and $E(e_i) = 0$ and that the variances of $\mathbf{u}$ and $\mathbf{e}$ are known non-singular variances matrices $\mathbf{A}$ and $\mathbf{R}$ and that u and e are uncorrelated with each other. The variance matrix of $\mathbf{y}$ is then $\mathbf{ZAZ}' + \mathbf{R} = \mathbf{V}$.

We note that if $\mathbf{A} = \mathbf{I}\sigma^2_b$ and $\mathbf{R} = \mathbf{I}\sigma^2$, then (10) is a matrix formulation of the mixed model version of (1) and β and $\mathbf{u}$ would then represent the herd-year-season and sire effects. In Section 2 we discussed methods of predicting the sire effects $\mathbf{u}$. Henderson (1973) considers the more general problem of predicting linear functions of the unknown β and $\mathbf{u}$. For instance we might be interested in predicting the merit of the daughter of sire j in an average herd-year-season, giving weight P_{ji} to the i*th* herd-year-season. That is, we wish to predict $w_j = \sum_i P_{ji}\beta_i + u_j$ or $\mathbf{w} = \mathbf{P}\beta + \mathbf{u}$. He suggests using a predictor $\hat{\mathbf{w}}$ for $\mathbf{w}$ where (i) $\hat{w}_i$ is a linear function of y, (ii) $\hat{w}_i$ is an unbiased estimator of w_i (i.e., $E(\hat{w}_i) = w_i$), (iii) the mean square error (i.e., $E(\hat{w}_i - w_i)^2$) is minimized. The expectations correspond to values in hypothetical repetitions of sampling and predicting and involve averaging over $\mathbf{e}$ and $\mathbf{u}$. Henderson shows that this best linear unbiased predictor (BLUP) of the elements of w satisfies

$$\hat{\mathbf{w}} = \mathbf{P}\hat{\beta} + \mathbf{AZ'V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}) \tag{11}$$

where $\hat{\beta}$ satisfies the least-squares solution

$$(\mathbf{X'V}^{-1}\mathbf{X})\beta = \mathbf{X'V}^{-1}\mathbf{y}. \tag{12}$$

We note that $\mathbf{AZ'}$ is the covariance of $\mathbf{u}$ with $\mathbf{y}$ and $\mathbf{V}$ the variance of $\mathbf{y}$ so that $\mathbf{AZ'V}^{-1}$ is informally cov($\mathbf{u}$,$\mathbf{y}$) "divided by" var($\mathbf{y}$), and so can be thought of as regression coefficients relating $\mathbf{y}$ (corrected for the fixed effects) to $\mathbf{u}$.

If β was known one could combine all the information available on relatives in a selection index (Smith 1936, Hazel 1943) in order to predict $\mathbf{u}$. The coefficients $\mathbf{b}_i$ of a predictor $\tilde{u}_i = \mathbf{b}_i'(\mathbf{y} - \mathbf{X}\beta)$ of u_i from a selection index satisfy $\mathbf{Vb}_i = \mathbf{ZAd}_i$ where $\mathbf{d}_i$, the vector of economic weights, is of length b with i*th* element one and all others zero. Hence $\mathbf{b}_i = \mathbf{V}^{-1}\mathbf{ZAd}_i$, $\tilde{u}_i = \mathbf{d}_i'\mathbf{AZ'V}^{-1}(\mathbf{y} - \mathbf{X}\beta)$, $\tilde{\mathbf{u}} = \mathbf{AZ'V}^{-1}(\mathbf{y} - \mathbf{X}\beta)$ and so a selection index predictor of $\mathbf{w}$ is then $\tilde{\mathbf{w}} = \mathbf{P}\beta + \mathbf{AZ'V}^{-1}(\mathbf{y} - \mathbf{X}\beta)$.

For example, for the mixed model (1) if a bull has n daughters $b_i = \sigma_b^2/(\sigma^2 + n\sigma_b^2) = 1/(n + \gamma)$ is the coefficient of each daughter's yield. Hence the coefficient of the bull's daughter's yield is $n/(n + \gamma)$ as used in (3). The BLUP predictor, $\hat{\mathbf{w}}$, is of the same form as the selection index predictor, $\tilde{\mathbf{w}}$, with the intuitively sensible replacement of the known value β in $\tilde{\mathbf{w}}$ by the least squares estimate $\hat{\beta}$ in $\hat{\mathbf{w}}$.

Formation of (11) and (12) requires the inversion of the $n \times n$ variance matrix $\mathbf{V}$ and Henderson in Henderson, Kempthorne, Searle and von Krosigk (1959) pointed out that this inversion can be avoided. The solution of (12), $\hat{\beta}$ satisfies the equations

$$\begin{bmatrix} \mathbf{X'R}^{-1}\mathbf{X} & \mathbf{X'R}^{-1}\mathbf{Z} \\ \mathbf{Z'R}^{-1}\mathbf{X} & \mathbf{Z'R}^{-1}\mathbf{Z} + \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R}^{-1}\mathbf{y} \\ \mathbf{Z'R}^{-1}\mathbf{y} \end{bmatrix}. \tag{13}$$

Because of the similarity of these equations to the least squares equations

$$\begin{bmatrix} \mathbf{X'R}^{-1}\mathbf{X} & \mathbf{X'R}^{-1}\mathbf{Z} \\ \mathbf{Z'R}^{-1}\mathbf{X} & \mathbf{Z'R}^{-1}\mathbf{Z} \end{bmatrix} \begin{bmatrix} \beta \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R}^{-1}\mathbf{y} \\ \mathbf{Z'R}^{-1}\mathbf{y} \end{bmatrix} \tag{14}$$

that arise when $\mathbf{u}$ are fixed, (13) has been called a mixed model equation. Henderson pointed out that $\hat{\mathbf{u}}$ in (13) is not just a computational artifact arising out of the solution for $\hat{\beta}$, and related $\hat{\mathbf{u}}$ (for his particular example) to Lush's (1949) most probable producing ability. Later Henderson (1963) showed that the predictor (11) can be written in terms of the solution of (12) as $\mathbf{w} = \mathbf{P}\hat{\beta} + \hat{\mathbf{u}}$, so that $\hat{\mathbf{u}}$ is the BLUP predictor of $\mathbf{u}$.

Several properties of least squares estimates of fixed effects have analogues in the prediction of random effects. For instance if

$$\begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}$$

is the inverse of the left hand side of (13) then $\mathbf{K'C}_{11}\mathbf{K}$ is the variance matrix of $\mathbf{K'}\hat{\beta}$ (the best linear unbiased estimate of $\mathbf{K'}\beta$). But $\mathbf{M'C}_{22}\mathbf{M}$ is not the variance matrix of $\mathbf{M'}\hat{\mathbf{u}}$ (the BLUP of $\mathbf{M'u}$) but the variance matrix of the prediction errors $\mathbf{M'}(\hat{\mathbf{u}} - \mathbf{u})$.

One difference between the least squares and mixed model equations is that the least squares equations (14) are not of full rank because adding together the first t equations of (14) gives the same equation as adding the last b equations of (14). Hence not all effects are estimable. However the presence of $\mathbf{A}^{-1}$ in the left hand side of the mixed model equations ensures that the mixed model equations are of full rank. Manipulation of (13) shows that $\mathbf{l'A}^{-1}\hat{\mathbf{u}} = 0$ where $\mathbf{l}$ is a $b \times 1$ vector of ones.

There are several other ways of deriving the mixed model equations, some more informative than others. Henderson (in Henderson *et al.* 1959) first derived (13) by maximizing for variation in β and $\mathbf{u}$ the joint density function of $\mathbf{y}$ and $\mathbf{u}$, $f(\mathbf{y},\mathbf{u})$, when $\mathbf{u}$ and $\mathbf{e}$ are normally distributed. This joint density function can arise by a Bayesian argument. One starts by assuming a model with $\mathbf{u}$ as fixed effects and $\mathbf{e}$ normally distributed and then also assume that there is prior information that $\mathbf{u}$ is normally distributed about a mean of $\mathbf{0}$ with variance $\mathbf{A}$ (for example Lindley and Smith 1972, Harville 1976). The posterior density distribution of β and $\mathbf{u}$ is then proportional to $f(\beta,\mathbf{u})$ and *a posteriori* β and $\mathbf{u}$ are normally distributed about a mean $\hat{\beta}$, $\mathbf{u}$. Finally, Dempfle (1977) uses two independent estimates of $\mathbf{u}$, one a prior estimate of zero and the other the least squares estimate. He combines these estimates using as weights their respective prediction variances and shows that the resulting estimate satisfies (13).

The development in this section is based on a simple full rank model with non-singular variance matrices. Extensions to include more factors or covariates and variates or traits are fairly obvious. Harville (1976) has recently considered a more general case, when the model is not necessarily of full rank and the variance matrices can be singular.

4. Selection

In the last section we discussed the evolution of methods required by the more effective selection of young bulls. In this section we discuss in more detail some of the effects of selection. The primary purpose of this section is to discuss a selection model of Henderson (1975). However, the example we use to motivate the discussion will be useful in later sections.

A convenient starting point is the paper of Henderson *et al.* (1959). They were interested in estimating genetic and environmental trends for dairy cattle in which culling had taken place. We will use a simple example where $n_1 + n_2$ cows have heifer lactations with mean $\bar{y}_1$ and then n_1 cows (with heifer mean $\bar{y}_{11}$) go on to have a second lactation in another environment. We suppose the first and second lactations, y_1 and y_2, are normally distributed about means μ_1 and μ_2 with variance V_{11} and V_{22} and there is a covariance between y_1 and y_2 of $V_{21}(= V_{12})$. If the cows that have second lactations are selected at random then the likelihood can easily be written down. Maximum likelihood (ML) estimates of the environmental effects then satisfy weighted least squares equations which can be written in the form of (12).

Now suppose that instead of random selection the cows that have second lactations are chosen on their phenotypic performance in their heifer lactation. Kempthorne and von Krosigk (in Henderson *et al.* 1959) suggest writing the log-likelihood as the sum of two independent parts, $\mathcal{L}_1$ the log-likelihood of the heifer records, y_1, and $\mathcal{L}_2$ the log-likelihood of the second record given the first record, $y_2 - V_{21}V_{11}^{-1}y_1$. They show that the form of the log-likelihood in this case is the same as when cows are selected at random to have second lactations. The resulting ML estimates of μ_1 and μ_2, $\hat{\mu}_1$ and $\hat{\mu}_2$, again satisfy (12). These estimates of μ_1 and μ_2 can be intuitively deduced by noting that $y_2 - V_{21}V^{-1}{}_{11}y_1$ has mean $\mu_2 - V_{21}V_{11}^{-1}\mu_1$ and variance $V_{22} - V_{11}^{-1}V_{12}$ and is independent of $y_1 - \mu_1$. Hence $\hat{\mu}_1 = y_1$ and $\hat{\mu}_2 = \bar{y}_2 - V_{21}V_{11}^{-1}(\bar{y}_{11} - \bar{y}_1)$. If there is selection on heifer yield then $\bar{y}_{11} > \bar{y}_1$ and the difference of the means of the lactations $\bar{y}_2 - \bar{y}_1$ will usually overestimate $\mu_2 - \mu_1$ and the difference of the means for the cows with two lactations $\bar{y}_2 = \bar{y}_{11}$ will usually underestimate $\mu_2 - \mu_1$.

This argument relies on selection, or culling, being on a trait in the model, but often selection is on an unknown trait correlated with the trait in the model. Henderson (1975) has introduced a conditional model to cover this situation but I find his approach hard to understand. He considers a model for $\mathbf{y}$ conditional on $\mathbf{r}$ of the form $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} +$

$\mathbf{BH}^{-1}\mathbf{r}+\mathbf{e}$ where var $(\mathbf{y})=\mathbf{R}+\mathbf{ZAZ}'-\mathbf{BH}^{-1}\mathbf{B}'$ and $\mathbf{B}\,(=\mathbf{ZB}_u+\mathbf{B}_e)$ and $\mathbf{H}$ are known matrices and $\mathbf{r}$ is an unknown vector. We note that the model for the second lactation given the first discussed above is of this form. Henderson derives BLUP equations for this general model. I find an informal derivation useful. A model of the form $\mathbf{y}=\mathbf{X}\beta+\mathbf{BH}^{-1}\mathbf{r}+\mathbf{Zu}+\mathbf{BH}^{-1}\mathbf{s}+\mathbf{e}$ where $\mathbf{s}$ is distributed with variance $-\mathbf{H}$ generates the required variances. Although, as $\mathbf{s}$ has negative variances it could perhaps be thought of as representing imaginary random variables. This model is of the form (10) with $\mathbf{X},\mathbf{Z},\beta,\mathbf{u}$ and $\mathbf{A}$ being replaced by $(\mathbf{X}\ \mathbf{BH}^{-1})$, $(\mathbf{Z}\ \mathbf{BH}^{-1})$, $(\beta\ \mathbf{r})$, $(\mathbf{U}\ \mathbf{s})$ and $(\mathbf{W}'\ \mathbf{E}')'$ where $\mathbf{W}=(\mathbf{A}\ \mathbf{O})$ and $\mathbf{E}=(\mathbf{O}\ -\mathbf{H})$, respectively. The BLUP equations are then

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{BH}^{-1} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{BH}^{-1} \\ \mathbf{H}^{-1}\mathbf{B}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{H}^{-1}\mathbf{B}'\mathbf{R}^{-1}\mathbf{BH}^{-1} & \mathbf{H}^{-1}\mathbf{B}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{H}^{-1}\mathbf{B}'\mathbf{R}^{-1}\mathbf{BH}^{-1} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{BH}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{A}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{BH}^{-1} \\ \mathbf{H}^{-1}\mathbf{B}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{H}^{-1}\mathbf{B}'\mathbf{R}^{-1}\mathbf{BH}^{-1} & \mathbf{H}^{-1}\mathbf{B}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{H}^{-1}\mathbf{B}'\mathbf{R}^{-1}\mathbf{BH}^{-1}-\mathbf{H}^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ \mathbf{r} \\ \mathbf{u} \\ \mathbf{s} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{H}^{-1}\mathbf{BR}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{H}^{-1}\mathbf{BR}^{-1}\mathbf{y} \end{bmatrix}$$

Manipulating the second and fourth equations shows $\mathbf{s}=\mathbf{o}$ and the resulting equations can be rewritten as

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{B} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{A}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{B} \\ \mathbf{B}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{B}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{B}'\mathbf{R}^{-1}\mathbf{B} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{u}} \\ \mathbf{H}^{-1}\hat{\mathbf{r}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{B}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}.$$

As $\mathbf{u}$ is distributed with mean $\mathbf{B}_u\mathbf{H}^{-1}\mathbf{r}$, conditional on $\mathbf{r}$, this suggests predicting $\mathbf{m}'\mathbf{u}$ by $\mathbf{m}'\hat{\mathbf{u}}+\mathbf{m}'\mathbf{B}_u\mathbf{H}^{-1}\hat{\mathbf{r}}$ which agrees with Henderson's more rigorous derivation. There is an analogy with least squares when the estimates of β under the model $E(\mathbf{y})=\mathbf{X}\beta$, $\text{var}(\mathbf{y})=\mathbf{V}$ are the same as under the model $E(\mathbf{y})=\mathbf{X}\beta$, $\text{var}(\mathbf{y})=\mathbf{V}+\mathbf{XSX}'$ where $\mathbf{S}$ is any symmetric matrix.

This is quite straightforward. Henderson goes on to consider the three cases when $\mathbf{r}$ the conditioning variable becomes $\mathbf{L}'\mathbf{y}$, $\mathbf{L}'\mathbf{u}$ or $\mathbf{L}'\mathbf{e}$. Why one would want to condition on $\mathbf{L}'\mathbf{y}$ I am not sure. Henderson argues that if in repeated sampling $E(\mathbf{L}'\mathbf{y}\neq\mathbf{L}'\mathbf{X}\beta$ then one should use the conditional model with $\mathbf{r}=\mathbf{L}'\mathbf{y}$. For instance, for the example above Henderson suggests using $\mathbf{L}'\mathbf{y}=\bar{y}_{11}-\bar{y}_{12}$ (where $\bar{y}_{12}$ is the mean of heifers with no second record). Using Kempthorne and von Krosigk's argument about the partition of the likelihood, this is equivalent to using all the information from $\mathfrak{L}_2$ (y_2 conditional on y_1) but only part of the information from $\mathfrak{L}_1$ (i.e., not using that provided by $\mathbf{L}'\mathbf{y}$). Why one cannot use the information provided by $\mathbf{L}'\mathbf{y}$ is not clear to me. Presumably at the earlier stage of just knowing y_1 and before knowing which heifers continue to a second lactation one is able to use $\mathbf{L}'\mathbf{y}$.

Another difficulty I have is that $\mathbf{L}$ depends on the selection carried out and so one can visualize it varying from one sample to another. Suppose in the example above we add different environmental effects to summer and winter calving heifers and selection of cows is on phenotypic performance regardless of when they calved (for example Hinks 1966). In this case I find it natural to take $\mathbf{L}'\mathbf{y}=(\bar{y}_{11s}-\bar{y}_{12s},\bar{y}_{11w}-\bar{y}_{12w})'$ where the suffices s and w refer to summer and winter calvers. Now in repeated sampling the numbers of spring cna winter calvers selected to have second lactations will vary and so will $\mathbf{L}$. Incidentally this example is a counterexample to Henderson's assertion (1973) that if $\mathbf{L}'\mathbf{X}$ is null, selection is *within* fixed effects, for in this example selection is *ignoring* fixed effects.

Just as in the unconditional model, one can derive the BLUP equations for the conditional model in several ways. For instance one can derive them using the joint density of $\mathbf{y}$ and $\mathbf{u}$ conditional on $\mathbf{r}$. Fimland (1975) has attempted to do this for some special values of $\mathbf{r}$ but his derivation is erroneous. He implicitly argues that the conditional variables $\mathbf{y} - \mathbf{B}\mathbf{H}^{-1}\mathbf{r}$ have the unconditional variance $\mathbf{R} + \mathbf{Z}\mathbf{A}\mathbf{Z}'$ when $\mathbf{r} = \mathbf{L}'\mathbf{y}$, $\mathbf{L}'\mathbf{u}$ or $\mathbf{L}'\mathbf{e}$ but there is no justification for this assumption.

5. *Relationships Between Animals*

The theory in Section 3 is developed using a general $\mathbf{A}$ matrix which allows relationships between sires to be taken into account. However until recently the knowledge, for instance that young bulls are the sons of proven bulls, has not usually been used and the $\mathbf{A}$ matrix has been replaced by a diagonal matrix. There are two reasons for this. One, if a bull has a reasonable number of daughters the extra information provided by the bull's sire is small. Information from the bull's sire is equivalent to $(4\text{-}h^2)/3h^2$ daughters (for instance 5 daughters if $h^2 = 0.25$) of the bull (Searle 1964*b*). Secondly, the *BLUP* equations (as set up in (13)) need the evaluation of $\mathbf{A}^{-1}$ and direct inversion is not practical.

Henderson (1976) has shown that if $\mathbf{A}$ includes additive covariances between individuals the inversion of $\mathbf{A}$ is feasible and that in fact it is easier to form $\mathbf{A}^{-1}$ than $\mathbf{A}$. This follows from the fact that $\mathbf{A}$ can be written as $\mathbf{T}\,\mathbf{D}\,\mathbf{T}'$ where $\mathbf{D}$ is a diagonal matrix and $\mathbf{T}$ is a lower triangular matrix with diagonal elements non-zero and i, jth element non-zero if the jth individual is an ancestor of the ith individual (Thompson 1977). The matrix $\mathbf{T}$ has a simple inverse with the diagonal elements non-zero and the i, jth elements non-zero if the jth individual is a parent of the ith individual. Hence $\mathbf{A}$ has a simple inverse.

It is interesting to note that the sire effect u_j can be written as an accumulation of independent terms from its ancestors and itself, i.e., $u_j = \sum_\ell T_{\ell j} u_\ell{}^*$, where the effects $u_j{}^*$ have variance matrix $\mathbf{D}$. First use of the $\mathbf{A}^{-1}$ matrix was in problems of sire evaluation but it has also been used in the parallel problem of evaluating cows. For example Everett, Henderson and Hanson (1977) use data from all lactations, the relationships between cows in a herd and predictions of sires' merit to give estimates of the transmitting ability of cows.

Use of the $\mathbf{A}$ matrix assumes certain relationships between covariances between relatives, for instance that the covariance between half-sibs is equal to the dam-daughter covariance and equal to twice the aunt-niece covariance. Whether the method should be used if these relationships do not hold is a debatable point. There is evidence that the dam-daughter covariance is greater than the half-sib covariance (Bradford and Van Vleck 1969). A related point is that if the young bulls used are sons of proven bulls the genetic variance among these bulls will be reduced by up to a third (Robertson 1977) and one might ask if a reduced value of heritability should be used. By analogy with the argument in the two lactation case in Section 4, when parameters in the unselected population are appropriate even when there is selection of dams, it seems that one should use the heritability in the unselected population if one uses the information provided by the sires of bulls. If the sires of bull information is not used then presumably it is better to use the heritability value in the selected population. Whether this is a mere theoretical quibble or of practical importance I do not know. The results of Sales and Hill (1976) and studies on alternative sire evaluation methods suggest that the ranking of sires with a reasonable number of daughters will be robust to changes in the value of heritability but that the estimates of genetic progress might be more affected. The ranking of dams should be between these extremes.

Theoretically the use of the relationship matrix can be thought of as an alternative to grouping sires according to time of entry into the stud. I think that use of the relationship

matrix should reduce the need for grouping sires. However, it is not entirely clear how group effects should be included in a model that uses relationships. One approach is to use a model of the form

$$y_{ijk} = \beta_i + u_j + G_{g(j)} + e_{ijk} \tag{15}$$

where the *jth* sire is in the *g(j)th* group. This model is used to evaluate bulls in the North East States of America. Another approach is to write u_j in terms of the independent effects u_j^* and use a model of the form

$$y_{ijk} = \beta_i + \sum_\ell T_{\ell j}(u_\ell^* + G_{g(\ell)}^*) + e_{ijk} \tag{16}$$

or equivalently, $y_{ijk} = \beta_i + u_j + \sum_\ell T_{\ell j} G_{g(\ell)}^* + e_{ijk}$.

For example if sire 2 is the son of sire 1 then (16) gives

$$y_{i1k} = \beta_i + u_1^* + G_1^* + e_{i1k'} \qquad y_{i2k} = \beta_i + \tfrac{1}{2}u_1^* + u_2^* + \tfrac{1}{2}G_1^* + G_2^* + e_{i2k}$$

or

$$y_{i1k} = \beta_i + u_1 + G_1^* + e_{i1k'} \qquad y_{i2k} = \beta_i + u_2 + \tfrac{1}{2}G_1^* + G_2^* + e_{i2k}.$$

This approach of accumulating group effects is also used by Henderson (1973). He suggests a maternal grandsire model including a term $(u_m + G_{g(m)})/2$ to represent the maternal grandsire effect of the *mth* grandsire. Models (15) and (16) will lead to different sire predictors but I have not seen the differences investigated. Intuitively I prefer (16) as (15) does not seem to use the ancestral group information. Model (16) will require more computational effort to fit than (15).

The USDA scheme of evaluating bulls does not directly use the information on a bull's sire, but the information is used indirectly in grouping the bulls according to their sire and maternal grandsire information. I would have thought that with this criterion of grouping bulls there could be quite a high probability of "mis-grouping" bulls. Of course the more daughters a bull has the less the emphasis on the group mean (Powell and Freeman 1974*b*).

6. *Genetic and Environmental Trend*

Most estimates of genetic and environmental trend have been using the method of Smith (1962) or some modification of his method. Essentially the regression of performance against time (β_{PT}) and of performance against time within sire (β_{PTS}) are used to give estimates of genetic (g) and environmental trend (e). Sometimes adjustments for the dam age or merit are made (for example, Powell and Freeman 1974*a*). These take the form of adjusting the regression coefficients β_{PT} and β_{PTS} using the regression of dam age or merit on time. It would seem more natural to me to consider the performance at time t as $p_t = \alpha + g(t - a) + et$ where a is an adjustment for dam age or merit so that regression of performance on time and adjusted time is used to estimate trends. This method assumes linearity of genetic and environmental trend and, except when adjustments are made for dam merit, makes no assumption on the value of heritability.

Recently the *BLUP* method has been taken to its logical conclusion and Hintz, Everett and Van Vleck (1978) have given estimates of genetic trend using *BLUP* predictors of bulls and cows. Van Vleck (1977) presents these results, plotting genetic merit for cows against year of first calving. I find it difficult to interpret such graphs because the changes in genetic merit over time are partly due to changes in selection pressure over time. One possibility, by analogy with selection experiments, is to plot genetic merit against selection differential. This might make it easier to interpret changes in genetic merit over time. It might also give an

empirical check on the value of heritability (compare with realized heritability estimation, Hill 1972). One might also compare the actual selection differentials for different sub-groups (young bulls, proven bulls etc.) with theoretical optimum selection differentials.

There are difficulties with this approach that I think are worth investigating. Consider the case when we have heifer records on $n_1 + n_2$ cows and on daughters of n_1 of these cows in a different environment. These data have the same structure as that in Section 3 and the *BLUP* estimate of difference in genetic merit between parents and offspring is $h^2(\bar{y}_{11} - \bar{y}_1)/2$. This is a simple multiple of the selection differential and does not depend on the offspring records. This is perhaps an extreme case in that the subclasses compared have different environmental effects. In practice one might not have as much confounding between genetic groups and environmental effects. Certainly this raises the question of whether the *BLUP* estimates of trend are just multiples of the selection differentials.

7. *Estimation of Variance Components*

The previous sections were concerned with prediction assuming that the variances were known. Often however we wish to estimate the variance parameters. In this section we discuss various methods used in sire evaluation and show how predicted values play an integral part of the more recent developments. We will again use the simple two-factor model (10) as an example but assume that $\mathbf{A}$ and $\mathbf{R}$ are diagonal matrices $\mathbf{I}\sigma_b^2$ and $\mathbf{I}\sigma^2$ respectively, i.e., $\mathbf{u}$ and $\mathbf{e}$ contain uncorrelated elements with homogeneous variances, and that σ_b^2 and σ^2 are unknown parameters.

Developments up to 1970 have been extensively discussed by Searle (1971). Most methods follow the simple prescription of finding two sums of squares and equating these to their expectation. One of the commonest methods (Henderson's method 3) is based on the least squares equations and consists of equating the residual sum of squares and the sum of squares for the random factor adjusted for the fixed factor to their expectation. The estimates of σ^2 and σ_b^2 then satisfy

$$\begin{aligned} \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{u}'\mathbf{Z}\mathbf{y} &= (n - t - b + 1)\sigma^2, \\ \mathbf{u}'\mathbf{Z}\mathbf{y} &= (b - 1)\sigma^2 + tr(\mathbf{Z}'\mathbf{Z} - \mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z})\sigma_b^2. \end{aligned} \tag{17}$$

The efficiency of this procedure is in general not known and naturally depends on the relative magnitude of σ^2/σ_b^2 and the degree of inbalance. This is rather a schizophrenic procedure in that a fixed effect model is first used to generate the sums of squares in (17) and then their expectation is found under a mixed model. Also if one accepts the mixed model one would like to use the estimates of fixed effects given by (12) in estimating the variance parameters. When there are more than two components to be estimated more sums of squares can be found than components to be estimated. It is not always obvious which sums of squares should be used.

This unease with established methods has partly been the impetus for research into other methods and we now outline some of the results. Harville (1977) has recently comprehensively reviewed these recent developments. Hartley and Rao (1967) discussed maximum likelihood estimation when $\mathbf{u}$ and $\mathbf{e}$ are normally distributed and showed that the ML equations satisfy

$$\begin{aligned} E &= \mathbf{y}'\mathbf{y} - \hat{\beta}\mathbf{X}'\mathbf{y} - \hat{\mathbf{u}}'\mathbf{Z}'\mathbf{y} = n\sigma^2 \\ B &= \hat{\mathbf{u}}'\hat{\mathbf{u}} = b\sigma_b^2 - tr(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1})^{-1}, \end{aligned} \tag{18}$$

where $\hat{\beta}$ and $\hat{\mathbf{u}}$ satisfy (12). These equations can be interpreted (Patterson and Thompson

1971) as equating a "residual" or error term, E, analogous to the residual sum of squares (compare (18) with (17)) and a sum of squares of predicted values, B, to their expectation. This expectation is taken assuming the fixed effects, β, are known and takes no account of the fact that the fixed effects β are estimated. Patterson and Thompson (1971) suggest one should, to avoid this difficulty, use only the likelihood of error contrasts, i.e., contrasts that have zero expectation, to estimate the variance parameters. The estimates of σ^2 and $\sigma_b{}^2$ (called restricted maximum likelihood (*REML*) estimates in Harville's (1977) review paper) then satisfy

$$E = (n - t)\sigma^2 \qquad \text{and} \qquad B = b\sigma_b{}^2 - tr(C_{22}). \tag{19}$$

The lefthand sides of (19) are as in (18) but the righthand sides of (19) can be interpreted as the expectation of the lefthand sides, taking into account the fact that β is estimated. We note that if $t = 0$ (19) reduces to (18).

Usually the variance estimates have to be found iteratively successively using estimates of σ^2 and $\sigma_b{}^2$ to solve (12) and then forming new estimates of σ^2 and $\sigma_b{}^2$ from $\hat{\beta}$ and $\hat{u}$. It is very appealing to use (19) or something similar (Henderson 1973, Harville 1977) to give new estimates but in practice this can be very slow to converge. As an example Schaeffer, Wilton and Thompson (1978) give a numerical example that takes over 50 iterations to converge satisfactorily using a scheme analogous to (19). Analytical results on the rate of convergence for the balanced one-way classification (Dempster, Laird and Rubin 1977) are in agreement with this finding and suggest that the rate of convergence depends on $\sigma_b{}^2/\sigma^2$ and goes to zero if the ML estimate of $\sigma_b{}^2$ is zero.

I have found that an iterative scheme based on Fisher's method of scoring converges much faster. First $\mathbf{F}_0$ and $\mathbf{d}_0$ are found, by replacing σ^2 and $\sigma_b{}^2$ by initial estimates in the following formulae (from Thompson 1973)

$$\mathbf{F} = \begin{pmatrix} tr(\mathbf{U}^2) & tr(\mathbf{U}\mathbf{C}_{22})/\sigma_b{}^2 \\ tr(\mathbf{U}\mathbf{C}_{22})/\sigma_b{}^2 & (n - t - b)/\sigma^4 + tr(\mathbf{C}_{22}{}^2)/\sigma_b{}^4\sigma^4 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} B/\sigma_b{}^4 \\ E/\sigma^4 - B/(\sigma^2\sigma_b{}^2) \end{pmatrix} \tag{20}$$

where $\mathbf{U} = \mathbf{I}\,\sigma_b{}^{-2} - (1/\sigma_b{}^4)\mathbf{C}_{22}$.

New estimates of σ^2 and $\sigma_b{}^2$ can then be found from

$$\mathbf{F}_0 \begin{pmatrix} \sigma_b{}^2 \\ \sigma^2 \end{pmatrix} = \mathbf{d}_0 \tag{21}$$

and the procedure repeated until the estimates converge to values, say, $\tilde{\sigma}^2$ and $\tilde{\sigma}_b{}^2$. The asymptotic variance matrix of $\tilde{\sigma}_b{}^2$ and $\tilde{\sigma}^2$ is then given by $2\tilde{\mathbf{F}}$. Again we note that for any choice of $\sigma_0{}^2$ and $\sigma_{b0}{}^2$ (21) can be interpreted as equating $\mathbf{d}_0$ to its expected value. Further, Lamotte (1973) has shown that when the true values of σ^2 and $\sigma_b{}^2$ are $\sigma_0{}^2$ and $\sigma_{b0}{}^2$, (21) gives the locally best (minimum variance) translation-invariant quadratic estimators of σ^2 and $\sigma_b{}^2$ when $\mathbf{u}$ and $\mathbf{e}$ are normally distributed.

Another justification of (19), making no assumption of normality, is provided by Rao's method of minimum norm quadratic unbiased estimation (MINQUE). Rao's estimators, for our example, are of the form $\hat{\sigma}^2 = \mathbf{y}'\mathbf{A}_1\mathbf{y}$ and $\sigma_b{}^2 = \mathbf{y}'\mathbf{A}_2\mathbf{y}$ where the matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ are symmetric and chosen to minimize the Euclidean norm (sum of squares of all the elements) of $\mathbf{A}_1\mathbf{U}$ and $\mathbf{A}_2\mathbf{U}$ where $\mathbf{U}$ is a positive definite matrix given by $\mathbf{U} = \alpha_1{}^2\mathbf{I} + \alpha_2{}^2\mathbf{Z}\mathbf{Z}'$. The minimization is subject to the condition that $E(\mathbf{y}'\mathbf{A}_1\mathbf{y}) = \sigma^2$ and $E(\mathbf{y}'\mathbf{A}_2\mathbf{y}) = \sigma_b{}^2$ which ensures unbiased estimation of σ^2 and $\sigma_b{}^2$. Then the MINQUE estimators of σ^2 and $\sigma_b{}^2$ satisfy (21) with $\mathbf{F}_0$ and $\mathbf{d}_0$ evaluated and with $\alpha_1{}^2$ and $\alpha_2{}^2$ replacing σ^2 and $\sigma_b{}^2$. In general the estimates of σ^2 and $\sigma_b{}^2$ depend on the choice of $\alpha_1{}^2$ and $\alpha_2{}^2$. Rao (1973) recommended that when possible

they should be chosen approximately proportional to σ^2 and σ_b^2. Whilst the MINQUE estimators have some intuitive appeal and reduce to other estimators when normality is assumed, it should be mentioned they do depend on a measure that is not accepted or understood by all statisticians.

One problem with the calculation of (19) and (20) is that they require the evaluation of the inverse of the lefthand side of (13) and in many problems this inversion might not be possible. Sometimes results about partitioned matrices reduce the problem to a manageable size. In other cases solutions to (13) are found by iteration. Perhaps the expansions for the inverse implicit in the iterative solution could be used to give acceptable approximations for the terms in (17) and (18). Harville (1977) has discussed other approximations to the ML method.

There is also the question of estimating variance parameters when there is selection. One problem is the definition of the variance parameters of interest. In the two-lactation example in Section 4 there are parameters in the unselected population V_{11}, V_{12} and V_{22}. There are also the values in the selected population. If in the heifer lactation the n_1 cows that have second lactations have variance M_{11} then, using the fact that $y_2 - V_{21}V_{11}^{-1}y_1$ has variance $V_{22} - V_{21}V_{11}^{-1}V_{12}$ and is independent of $y_1 - \mu_1$, it can be seen that in the selected population

$$\text{var}\,(y_2) = V_{22} - V_{21}V_{11}^{-1}(V_{11} - M_{11})V_{11}^{-1}V_{12} \quad \text{and} \quad \text{cov}\,(y_2, y_1) = V_{21}V_{11}^{-1}M_{11}. \tag{22}$$

There have been essentially three approaches to the estimation of variances in selected populations. One is to forget about the selection and just use standard analysis of variance, regression or correlation methods (for example, Butcher and Legates 1976). A second is to use standard methods but to correct for the selection using formulae derived by similar arguments to those used in (22) (for instance, Barker and Robertson 1966) or using formulae derived by simulation (Van Vleck 1968). A third approach (Curnow 1961, Thompson 1973) is to use a ML approach that essentially estimates the parameters in the unselected population. Curnow (1961) has considered the two-lactation example, with $V_{11} = V_{22}$, in detail and points out that there are three pieces of information on the variance parameters (*i*) from y_1, (*ii*) from regression of y_2 on y_1 and (*iii*) the deviation of y_2 from the regression on y_1. He mentions methods that use part of this information and shows that ML estimation uses all the information. Since the likelihood in this case is the same as when there is no selection, the REML estimates of V_{11} and V_{12} satisfy equations of the form (18) and it can be shown that these are unbiased estimating equations even if there is selection of cows on first lactation. However the iterative scheme depends on the expected value of the information matrix and this should be evaluated conditional on y_1 (Curnow 1961). This leads to simple modifications to (20) (Thompson 1973).

8. *Generalization of Linear Models*

This section has been motivated by the analysis of survival or culling data. Many analyses of this type of data start by first representing survival up to a particular stage by a variate coded as 1 or 0, depending on whether or not the animal survives up to that stage and then linear models are used to interpret these variates. I have some unease at using linear models for these dichotomous traits. Firstly, in many other applications to dichotomous variables logistic and probit (fixed effect) models have been used successfully. Cox (1970) discusses the advantages of such models. Secondly, the intra-class correlation on the (0, 1) scale depends on the proportion of the 1 class (i.e., proportion surviving). Often an assumption of an underlying continuous scale has been used (e.g., Falconer 1965, Curnow and Smith 1975). Individuals are scored one if they exceed a certain threshold value of liability. This model can

lead to a frequency independent estimate of correlation by a simple transformation of the (0, 1) scale correlation (Robertson and Lerner 1949). The related problem of evaluating sires under this model does not seem to have been investigated.

One approach is to try to parallel the development in Section 3 for this non-linear model. A suitable start is to assume that each element of $\mathbf{y}$ is binomially distributed with mean value $\Phi(\mathbf{X}\beta + \mathbf{Zu})$ i.e. associated with each animal is a liability variable normally distributed about a mean zero with a threshold depending on the effects β and the sire values $\mathbf{u}$. This is a generalized linear model (Nelder and Wedderburn 1972) and ML estimates of β and $\mathbf{u}$ satisfy equations of the form (13) and $\mathbf{y}$ is replaced by $\mathbf{X}\beta + \mathbf{Zu} + [\mathbf{y} - \Phi(\mathbf{X}\beta + \mathbf{Zu})]\,\mathbf{W}$, and $\mathbf{W}$ and $\mathbf{R}$ are diagonal matrices functions of $\mathbf{X}\beta + \mathbf{Zu}$. These equations are non-linear in $\mathbf{y}$ and have to be solved iteratively.

This fixed effect model assumes the underlying liability variables to be independently normally distributed. Instead of assuming that $\mathbf{u}$ are fixed effects, we now assume they are normally distributed with variance $\mathbf{A}$. This corresponds to assuming that the underlying liability variates are correlated for daughters of related bulls. Intuitively, this suggests using (3) with $\mathbf{R}$ and $\mathbf{y}$ modified as in the fixed effect model to give β and predictions of $\hat{\mathbf{u}}$. There is, I think, justification for this in the work of Dempster *et al.* (1977). Leonard (1972) has suggested similar estimators for a logistic model with binomial data. Again, if $\mathbf{A} = \mathbf{I}\sigma_b^2$ and we need to estimate σ_b^2, one might consider using $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ and something similar to (10) to estimate σ_b^2. (In this model there is no residual variance to estimate as we effectively scale the underlying variates to have residual variance 1). Leonard (1972) uses $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ and something simpler than (10). His formulae do not take account that β is estimated and I have found in several numerical examples that his method has given zero estimates for σ_b^2 where I would have expected non-zero estimates.

I think this is a potentially interesting area. The methods have been outlined using the binomial distribution and assuming a normal distribution for $\mathbf{u}$. If they are feasible they could be presumably easily generalized to deal with other distributions in the exponential family (e.g., Poisson or gamma distributions). Another possible extension is to consider cases when a transformation is needed to make $\mathbf{u}$ normal.

Acknowledgments

I am grateful to Dr. W. G. Hill and two referees for several useful comments.

Résumé

On passe en revue les développements récents dans le domaine des procédures d'évaluation de géniteurs, et on indique des relations entre plusieurs d'entre elles. L'approche du "meilleur prédicteur linéaire sans biais" (BLUP) est discutée. Ceci donne un cadre commode pour discuter des développements récents en estimation des composantes de la variance. On discute aussi d'autres sujets comme la sélection, et l'utilisation de relations entre animaux. On indique les domaines où des recherches supplémentaires sont nécessaires, comme l'estimation d'une tendance génétique et les méthodes d'évaluation pour les modéles non-linéaires.

References

Bar-Anan, R. (1976). The modified cumulative difference. *Proceedings of the 27th Meeting EAAP*, Zurich, Switzerland.

Bar-Anan, R. and Sacks, J. M. (1974). Sire evaluation and estimation of genetic gain in Israeli dairy herds. *Animal Production 18*, 59–66.

Barker, J. S. F. and Robertson, A. (1966). Genetic and phenotypic parameters for the first three lactations in Friesian cows. *Animal Production 8,* 221–240.

Bradford, G. E. and Van Vleck, L. D. (1964). Heritability in relation to selection differential in cattle. *Genetics 49,* 819–827.

Butcher, K. R. and Legates, J. E. (1976). Estimating son's progeny test from his pedigree information. *Journal of Dairy Science 59,* 137–152.

Cox, D. R. (1970). *The Analysis of Binary Data.* Methuen, London.

Cunningham, E. P. (1965). The evaluation of sires from progeny test data. *Animal Production 7,* 221–231.

Curnow, R. N. (1961). The estimation of repeatability and heritability from records subject to culling. *Biometrics 17,* 553–566.

Curnow, R. N. and Smith, C. (1975). Multifactorial models for familial diseases in man (with discussion). *Journal of the Royal Statistical Society, Series A 138,* 131–169.

Dempfle, L. (1977). Relation entre BLUP (Best Linear Unbiased Prediction) et estimateurs Bayesiens. *Annales de Génétique et de Sélection Animale 9,* 27–32.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B 39,* 1–38.

Dickinson, F. N., Norman, H. D., Powell, R. L., Waite, L. G. and McDaniel, B. T. (1976). Procedures used to calculate the USDA-DHIA modified contemporary comparison. *USDA Production Report No. 15,* 18–34.

Everett, R. W., Henderson, C. R. and Hanson, C. F. (1977). *The Northeast Cow ETA Report.* Department of Animal Science, Cornell University, Ithaca, New York.

Falconer, D. S. (1965). The inheritance of liability to certain diseases estimated from the incidence among relatives. *Annals of Human Genetics 29,* 51–76.

Fimland, E. A. (1975). Estimation of sire's breeding value. IV. Maximum likelihood derivation of the predictor and the estimator of a general mixed model given some type of selection of a correlated random variable. *Zeit. Tierz. Zuchtungsbiol. 93,* 131–135.

Hartley, H. O. and Rao, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika 54,* 93–108.

Harville, D. A. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. *Annals of Statistics 4,* 384–395.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association 72,* 320–338.

Hazel, L. N. (1943). The genetic basis for constructing selection indexes. *Genetics 28,* 476–490.

Henderson, C. R. (1963). Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding.* W. D. Hanson and H. F. Robinson, (eds.) Publication 982, National Academy of Sciences, National Research Council, Washington DC, 141–163.

Henderson, C. R. (1973). Sire evaluation and genetic trends. In *Proceedings of Animal Breeding Genetics Symposium, Blacksburg, Virginia.* American Society of Animal Science, Champaign, Illinois, 10–41.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics 31,* 423–449.

Henderson, C. R. (1976). A simple method for computing the inverse of a numerator matrix used in prediction of breeding values. *Biometrics 32,* 69–84.

Henderson, C. R. (1978). Undesirable properties of regressed least squares prediction of breeding values. *Journal of Dairy Science 61,* 114–120.

Henderson, C. R., Kempthorne, O., Searle, S. R. and Von Krosigk, C. N. (1959). Estimation of environmental and genetic trends from records subject to culling. *Biometrics 13,* 192–218.

Hill, W. G. (1972). Design and efficiency of selection experiments. I. Divergent selection. II. Selection in one direction. *Biometrics 29,* 747–770.

Hinks, C. J. M. (1966). Selection practices in dairy herds. 1. First lactation performance and survival to the second lactation. *Animal Production 8,* 467–480.

Hintz, R. L., Everett, R. W. and Van Vleck, L. D. (1978). Estimation of genetic trends from cow and sire evaluations. *Journal of Dairy Science 61,* 607–613.

LaMotte, L. R. (1973). Quadratic estimation on variance components. *Biometrics 29,* 311–330.

Leonard, T. (1972). Bayesian methods for binomial data. *Biometrika 51,* 581–589.

Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B 34,* 1–18.

Lush, J. L. (1949). *Animal Breeding Plans,* (3rd ed.) Iowa State College Press, Ames, Iowa.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A 135,* 370–384.
Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika 58,* 545–554.
Powell, R. L. and Freeman, A. E. (1974*a*). Genetic trend estimators. *Journal of Dairy Science 57,* 1067–1075.
Powell, R. L. and Freeman, A. E. (1974*b*). Estimators of sire merit. *Journal of Dairy Science 57,* 1228–1233.
Rao, C. R. (1973). *Linear Statistical Inference and its Applications,* (2nd ed.). Wiley, New York.
Robertson, A. (1966). A mathematical model of the culling process in dairy cattle. *Animal Production 8,* 95–108.
Robertson, A. (1977). The effect of selection in the estimation of genetic parameters. *Zeit. Tierz. Zuchtungsbiol. 94,* 131–135.
Robertson, A. and Lerner, I. M. (1949). The heritability of all-or-none traits: viability of poultry. *Genetics 34,* 395–411.
Robertson, A. and Rendel, J. M. (1954). The performance of heifers got by artificial insemination. *Journal of Agricultural Science, Cambridge 44,* 184–192.
Sales, J. and Hill, W. G. (1976). Effect of sampling errors on efficiency of selection indices. 1. Use of information from relatives for single trait improvement. *Animal Production 22,* 1–17.
Schaeffer, L. R., Wilton, J. W. and Thompson, R. (1978). Simultaneous estimation of variance and covariance components from multitrait mixed model equations. *Biometrics 34,* 199–208.
Searle, S. R. (1964*a*). Review of sire-proving methods in New Zealand, Great Britain and New York State. *Journal of Dairy Science 47,* 402–412.
Searle, S. R. (1964*b*). Progeny tests of sire and son. *Journal of Dairy Science 47,* 414–420.
Searle, S. R. (1971). *Linear Models.* Wiley, London.
Smith, C. (1962). Estimation of genetic change in farm livestock using field records. *Animal Production 4,* 239–251.
Smith, H. F. (1936). A discriminant function for plant selection. *Annals of Eugenics 7,* 240–250.
Thompson, R. (1973). The estimation of variance and covariance components when records are subject to culling. *Biometrics 29,* 527–550.
Thompson, R. (1976). Relationship between the cumulative difference and best linear unbiased predictor methods of evaluating bulls. *Animal Production 23,* 15–24.
Thompson, R. (1977). The estimation of heritability with unbalanced data. II. Data available on more than two generations. *Biometrics 33,* 497–504.
Van Vleck, L. D. (1968). Selection bias in estimation of the genetic correlation. *Biometrics 24,* 951–962.
Van Vleck, L. D. (1977). Theoretical and actual genetic progress in dairy cattle. In *Proceedings of the International Conference on Qualitative Genetics.* E. Pollak, O. Kempthorne and T. B. Bailey (eds.) Iowa State University Press, Ames, Iowa, 543–568.

Reprinted from
BIOMETRICS COPYRIGHT © 1979
THE BIOMETRIC SOCIETY, Vol. 35, No. 1, March 1979

26

Estimation of realised heritability in a selected population using mixed model methods

by

R. Thompson

Génét. Sél. Evol., 1986, **18** (4), 475-484

Estimation of realized heritability in a selected population using mixed model methods

R. THOMPSON

Animal Breeding Research Organisation, Edinburgh, Scotland

Summary

The use of mixed model methodology to estimate selection response and realized heritability from selection experiments with no controls is investigated. It has been suggested that the regression of predicted genetic worth on cumulative selection differential gives an estimate of heritability in a selected population. An assumed value of heritability is used to predict the genetic worth. It is shown for 2 simple designs, using pedigree information on one sex with both discrete and overlapping generations, that the predicted values depend crucially on the assumed value of heritability and not on the heritability in the population. Hence the regression estimator does not give an estimate of heritability in the selected population.

Key words : Realized heritability, mixed models.

Résumé

Application des méthodes du modèle mixte à l'estimation de l'héritabilité réalisée dans une population soumise à sélection

Cette étude concerne l'application de la méthodologie du modèle mixte à l'estimation de la réponse à la sélection et de l'héritabilité réalisée dans des expériences de sélection sans témoin. Il a été suggéré que la régression de la valeur génétique prédite sur la différentielle cumulée de sélection fournit une estimation de l'héritabilité dans une population soumise à sélection. Une valeur supposée de l'héritabilité est utilisée pour prédire la valeur génétique. On montre, dans 2 dispositifs simples utilisant l'information sur les apparentés dans un seul sexe, avec générations séparées ou chevauchantes, que les valeurs prédites dépendent de façon critique de la valeur supposée de l'héritabilité et non de l'héritabilité dans la population. Par suite, l'estimateur de la régression ne fournit pas une estimation de l'héritabilité dans la population soumise à sélection.

Mots clés : Héritabilité réalisée, modèle mixte.

I. Introduction

In experiments to evaluate the response to selection there is often a need to disentangle genetic trend from environmental effects. Two possibilities are to use divergent selection schemes (HILL, 1972a) or to use a control group (HILL, 1972b).

These designs allow regression of response on selection differential to give estimates of realised heritability (FALCONER, 1981). BLAIR and POLLAK (1984) investigated a third possibility of using mixed model methodology on a selected population to estimate genetic response. Mixed model methodology was first suggested as a means of separating genetic from environmental trends in dairy cattle records subject to culling (HENDERSON *et al.*, 1959). It was later discovered that this technique was a more powerful concept userful in problems of prediction of breeding values corrected for fixed effects (HENDERSON, 1973) and estimation of variance components by maximum likelihood (HARVILLE, 1977).

BLAIR & POLLAK (1984), by analogy with selection experiments, suggested using the regression of predicted yearly genetic means on the cumulative selection differential, b_{BP}, as an estimate of realised heritability. For one particular sheep selection experiment they state the standard error of b_{BP} is about the same as that for an estimate based on using both selected and control flock data. If true, this would be a remarkable result showing that selection experiments could be more efficient without control lines than with control lines. BLAIR & POLLAK (1984) noted that the results and conclusions may be influenced by the heritability value used in the prediction process. It seemed useful to quantify the extent of this influence. Partly to see if it was merely a numerical artefact, and partly because in one simple case, briefly discussed by THOMPSON, 1979, b_{BP} is exactly the value of heritability used to predict the breeding values.

By algebraically considering 2 simple designs it is shown that the regression coefficient does not give an estimate of heritability in the selected population. The designs considered in detail are partly motivated by actual selection experiments in this institute (PURSER, 1980), comments by BLAIR & POLLAK (1984) and last, but not least, algebraic simplicity. In both designs a pool of dams of constant genetic merit is assumed and pedigree information on the female side is ignored (a common occurence in dairy sire evaluation). Design I is a design where in each of T years sn males are measured and after the first year n sons of each of s sires are measured. In order to reduce genetic drift, suppose there is within family selection on the basis of the measured trait so that only one son of each sire is used as a sire. This design has no overlap between generations.

In design II suppose again sn males are measured in year 1. Then suppose s sires are selected using the measured trait and they each have n sons in year 2 and in year 3. One son is selected from each of the s sire families in year 2 and has n sons in year 3. There is now overlap is year 3 with offspring from sires of age 1 and 2.

II. Analysis

A. *Design I*

The observation and predicted additive genetic value for the ith animal in year 1 will be written as y_i and s_i ($i = 1, \ldots, sn$) and let $i = 1, \ldots, s$ represent sires that have offspring, also y_{tij} and s_{tij} represent the measurement and predicted value for the jth descendant of sire i in year t, and let $j = 1$ denote the individuals that have offspring.

Then suppose

$$\sum_{i=1}^{sn} y_{1i} = sn\bar{y}_1, \quad \sum_{i=1}^{n} y_{il} = s\bar{y}^*_1, \quad \sum_{i,j} y_{tij} = sn\bar{y}_t, \quad \sum_{i} y_{til} = s\bar{y}^*_t$$

$$\sum_{i=1}^{sn} s_{1i} = sn\bar{s}_1, \quad \sum_{i=1}^{n} s_{il} = s\bar{s}^*_1, \quad \sum_{j} s_{tij} = n\bar{s}_{ti}, \quad \sum_{i} s_{ti} = s\bar{s}_t, \quad \sum_{i} s_{til} = \bar{s}^*_t$$

It is assumed that there are fixed effects, m_t, associated with measurements on the t-th year and a residual variance σ^2_e associated with each observation. When there is no selection the genetic covariances between sires can be derived from the coefficients of parentage and the additive genetic variance, σ^2_A, (KEMPTHORNE, 1957) and this variance matrix will be denoted by $A\sigma^2_A$. It is well known that the genetic variances change with selection, but if σ^2_{AS} is thought of as the additive genetic variance in the base population before selection and selection is on traits included in the model, then a conditional argument can be used to show that operationally one can use $A\sigma^2_{AS}$ as the genetic variance matrix when estimating fixed effects (HENDERSON *et al.*, 1959) and when estimating σ^2_{AS} (CURNOW, 1961 ; THOMPSON, 1979).

In the appendix, mixed model equations are given and manipulated to show that estimators of genetic merit and year effects for this model are :

$$\bar{s}_1 = 0, \; m_1 = y_1$$
$$\bar{s}_2 = (1/2)\, h^2\, (\bar{y}^*_1 - \bar{y}_1), \; m_2 = \bar{y}_2 - (1/2)\, h^2\, (\bar{y}^*_1 - \bar{y}_1)$$
$$\bar{s}_{t+1} = (1/2)\, h^2_w\, (\bar{y}^*_t - \bar{y}_t) + 1/2\, \bar{s}_t, \; m_t = \bar{y}_t - \bar{s}_t$$

where h^2 is a prior estimate of heritability and $h^2_w = 3h^2/(4 - h^2)$ is a within half-sib family estimate of heritability.

BLAIR & POLLAK (1984) suggest regressing $\bar{s}_t$ on the cumulative selection differential. For this design the cumulative selection differential, CSD_t, at the end of year t satisfies $CSD_t = (1/2)\, CSD_{t-1} + (1/2)\, (\bar{y}^*_t - \bar{y}_t)$ with $CSD_o = 0$.

Then :

$$CSD_t - (1/2)\, CSD_{t-1} = (1/2)\, (\bar{y}^*_t - \bar{y}_t)$$
$$\bar{s}_{t+1} = h^2_w\, (CSD_t - (1/2)\, CSD_{t-1}) + (1/2)\, \bar{s}_t$$
$$\bar{s}_t = h^2_w\, (CSD_{t-1} - (1/2)\, CSD_{t-2}) + (1/2)\, \bar{s}_{t-1}$$
$$\bar{s}_3 = h^2_w\, (CSD_2 - (1/2)\, CSD_1) + (1/2)\, \bar{s}_2$$
$$\bar{s}_{t+1} = h^2_w\, (CSD_t - (1/2^{t-1})\, CSD_1) + (1/2^{t-1})\, \bar{s}_2$$
$$\bar{s}_{t+1} = h^2_w\, CSD_t + (1/2^{t-1})\, [h^2 - h^2_w]\, CSD_1$$

The predicted mean genetic merit is a multiple of the cumulative selection differential plus a correction term for the difference in heritabilities in the first and succeeding years that halves each year. The regression of $\bar{s}_{t+1}$ on CSD_t gives a regression coefficient lying between h^2_w and h^2 and tending to h^2_w as t increases.

B. *Design II*

A slight extension of the notation is needed to deal with this design because sires have sons in 2 years. Let y_{13ij} be measurements in year 3 on sons of males in year 1, with a similar definition of sire effects.

Also let

$\sum_{ij} y_{13ij} = sn\bar{y}_{13}$, $\sum_{j} s_{13ij} = n\bar{s}_{13i}$, $\sum_{i} \bar{s}_{13i} = s\bar{s}_{13}$, and

$$z(h^2) = (\bar{y}_{13} - \bar{y}_1) + (1/2)\, h_w^2\, (\bar{y}_2^* - \bar{y}_2) - (1/4)\, h^2\, (\bar{y}_1^* - \bar{y}_1) \quad (1)$$

It can be shown from the results in the appendix, that estimates of genetic merit using an assumed value of heritability, h^2, are

$$\bar{s}_1 = q_1 z(h^2) \quad (2)$$

$$\bar{s}_2 = (1/2)\, h^2\, (\bar{y}_1^* - \bar{y}_1) + q_2 z(h^2) \quad (3)$$

$$(\bar{s}_{13} + \bar{s}_3)/2 = (3/8)\, h^2\, (\bar{y}_1^* - \bar{y}_1) + (1/4)\, h_w^2\, (\bar{y}_2^* - \bar{y}_2) + q_3 z(h^2) \quad (4)$$

The terms, q_1, q_2, q_3 and can be found from functions of n and h^2. In table 1 are given values of q_1, q_2 and q_3 for various values of n and h^2 showing that q_1, q_2, q_3 increase as h^2 increases and that q_1 and q_3 decrease and q_2 increases as n increases.

TABLE 1

Values of q_1, q_2 and q_3 for various values of assumed heritability (h^2) and family size (n)

h^2	n	q_1	q_2	q_3
0.1	1	0.013	− 0.013	0.000
	2	0.012	− 0.007	0.000
	5	0.012	0.100	0.000
	10	0.011	0.035	− 0.002
	20	0.010	0.078	− 0.002
	30	0.009	0.114	− 0.002
0.3	1	0.039	− 0.039	0.000
	2	0.038	− 0.025	− 0.001
	5	0.035	0.014	− 0.003
	10	0.031	0.066	− 0.006
	20	0.025	0.142	− 0.010
	30	0.021	0.194	− 0.013
0.5	1	0.067	− 0.067	0.000
	2	0.064	− 0.048	− 0.002
	5	0.058	0.000	− 0.006
	10	0.050	0.063	− 0.012
	20	0.039	0.147	− 0.020
	30	0.032	0.201	− 0.025

The cumulative selection differentials in this case are $CSD_1 = (1/2)\, (\bar{y}_1^* - \bar{y}_1)$ and $CSD_2 = (3/8)\, (\bar{y}_1^* - \bar{y}_1) + (1/4)\, (\bar{y}_2^* - \bar{y}_2)$ so that $\bar{s}_{t+1} = h_w^2\, CSD_t + (1/2^{t-1})\, (h^2 - h_w^2)\, CSD_1 + q_{t+1}\, z(h^2)$, the same form as for design I with the addition of an extra term $q_{t+1}\, z(h^2)$.

In order to interpret the $z(h^2)$ term, it is seen, by considering the year effects estimators from design I, that there are in design II two estimates of m_3 readily available i.e. $m_3 = \bar{y}_3 - (1/2)\, h_w^2\, (\bar{y}_2^* - \bar{y}_2) - (1/4)\, h^2\, (\bar{y}_1^* - \bar{y}_1)$ and $m_3 = \bar{y}_{13} - (1/2)\, h^2\, (\bar{y}_1^* - \bar{y}_1)$. The discrepancy between these 2 values is $z(h^2)$. This is used in the mixed model approach to provide information on m_1 and m_2, and q_1 and q_2 can be

interpreted as measures of the information that z (h^2) provides on m_1 and m_2. Alternatively consideration of repeat-mating designs (for example, GIESBRECHT & KEMPTHORNE, 1965) suggest estimating $h_s^2 = \sigma_{AS}^2/(\sigma_e^2 + \sigma_{AS}^2)$ the heritability in the population by choosing h_s^2 so that z (h_s^2) = 0.

As the expected value of $\bar{y}_3 - \bar{y}_{13}$ is $1/2\, h_{sw}^2\, (\bar{y}_2^* - \bar{y}_2) - (1/4)\, h_s^2\, (\bar{y}_1^* - \bar{y}_1)$ the expected value of b_{BP}, E (b_{BP}) is a function of h^2, n, h_s^2 and the selection differentials. When $(\bar{y}_1^* - \bar{y}_1) = (\bar{y}_2^* - \bar{y}_2)$ then n and h_s^2 have little effect on E (b_{BP}). For instance for n = 30 and $h_s^2 = 0{,}1$ then E (b_{BP}) = 0.094, 0.291 and 0.493 when h^2 = 0.1, 0.3 and 0.5 and when n = 30 and $h_s^2 = 0.5$ then E (b_{BP}) = 0.088, 0.280 and 0.483. Again showing the crucial dependence of b_{BP} on the assumed value of heritability h^2 and not on the population value h_s^2.

III. Discussion

In 2 simple designs it has been shown that b_{BP} does not estimate heritability in the selected population. This should not be surprising in design I because of the confounding between years and generations. It is worrying in design II when a natural estimator of heritability is available.

Actual selection experiments, including the one considered by BLAIR & POLLAK (1984), are often more complicated than these 2 designs. For instance (i) mass or index selection could be carried out, (ii) measurements and pedigrees on females might be available, (iii) there could be more overlapping of generations, (iv) other effects such as sage of dam, partially confounded with generations need to be estimated, (v) it is rare to have equal family sizes. To take account of (i) one could try to explain s_t in terms of selection differentials within and between families. But if the phenotypic selectional differentials were used one would expect b_{BP} to be larger than h_w^2, but not as large as h^2. The actual magnitude depends on h_s^2 and the actual selection scheme. The major consequence of (ii) would be to reduce h_w^2 to within full-sib heritability $h^2/(2 - h^2)$. With (iii) the definition of the cumulative selection differential needs more care and there is the need to take account of the cumulative selection differential in the contemporaries (PURSER, 1980 ; JAMES, 1977). Both (iv) and (v) add some complexity to the analysis. None of these reasons suggest that b_{BP} will ever be a reasonable estimator of heritability from selection experiments without control.

As the estimated means $\bar{s}_t$ are derived from selection differentials they are not observed responses. Therefore the variances of $\bar{s}_t$ are not expressible in terms of the drift variances HILL (1972a) derives for observed responses as BLAIR & POLLAK (1984) assume.

Obviously when the value of h^2 used in predicting $\bar{s}_t$ is the value in the population then the calculation of $\bar{s}_t$ and b_{BP} can be useful as a monitoring device for the selection scheme and can be thought of as a sophisticated version of the predicted response h^2CSD rather than as a measured response to selection. However in selection experiments there will usually be the need to generate internally some evidence or tests for the value of parameters in the model including heritability. Just because some prediction of $\bar{s}_t$ is available from selection experiments without controls using mixed model methods does not seem to me sufficient grounds for recommending the use of such designs.

Of course there are other methods of estimating heritability and some of these are related to equating sums of squares of predicted values to their expectation (THOMPSON, 1977 ; SORENSEN & KENNEDY, 1984). However just because estimates are available does not imply that designs without controls are particularly efficient.

As a simple example consider 2 designs for 2 generations with N males measured in the first generation. In the first design offspring are raised from the best 2n males and heritability estimated by regression of offspring on parent. In the second, n males are chosen at random (a control) and the best n from the remaining N. In this design 2 natural estimates are possible one by comparing the response and selection differential and another by regression of offspring on parent. The variances of the 3 estimators are then inversely proportional to $(1 - i\,(i - x))$, $i'^2/2$ and $1 - i'\,(i' - x')/2 + i'^2/2$ (for example HILL (1970)) where x and z are the truncation point and ordinate for a normal distribution with a proportion $p - 2n/N$ truncated and $i = z/p$ and x', z', i' are the corresponding values for $p' = n/(N - n)$. For example with $n = 10$ and $N = 100$ then relative to the variance estimator in the first design, the two estimators in the second design variance $0.68/4.57 = 0.15$ and $0.49/4.57 = 0.11$ showing that the design with a control provides almost 10 times as much information on heritability as the design without a control.

Received January 2, 1986.
Accepted May 6, 1986.

References

BLAIR H.T., POLLAK E.J., 1984. Estimation of genetic trend in a selected population with and without the use of a control population. *J. Anim. Sci.*, **58**, 878-886.

CURNOW R.N., 1961. The estimation of repeatability and heritability from records subject to culling. *Biometrics*, **17**, 553-556.

FALCONER D.S., 1981. *Introduction to Genetics Statistics* (2nd ed.). 340 pp., Longman, London.

GIESBRECHT F., KEMPTHORNE O., 1965. Estimation of a repeat mating design for estimating environmental and genetic trends. *Biometrics*, **21**, 63-85.

HARVILLE D.A., 1977. Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.*, **72**, 320-338.

HENDERSON C.R., 1973. Sire evaluation and genetic trends. In *Proceeding of the Animal Breeding and Genetics Symposium in Honor of Dr. J.L. Lush*, 111, 10-41. ASAS and ADSA, Champaign.

HENDERSON C.R., KEMPTHORNE O., SEARLE S.R., VON KROSIGK C.N., 1959. Estimation of environmental and genetic trends from records subject to culling. *Biometrics*, **13**, 192-218.

HILL W.G., 1970. Design of experiments to estimate heritability by regression of offspring on selected parents. *Biometrics*, **26**, 566-571.

HILL W.G., 1972a. Estimation of realized heritabilities from selected experiments. I. Divergent selection. *Biometrics*, **28**, 747-765.

HILL W.G., 1972b. Estimation of realized heritabilities from selection experiments. II. Selection in one direction. *Biometrics*, **28**, 767-780.

JAMES J.W., 1977. A note on selection differential and generation length when generations overlap. *Anim. Prod.*, **24**, 109-112.

KEMPTHORNE O., 1957. An introduction to genetic statistics. 545 pp., Wiley, New York.

PURSER A.E., 1980. Comparison of expected and realised responses in three sheep selection experiments. *In* ROBERTSON A. (ed.), *« Selection Experiments in Laboratory and Domestic Animals »*, 21-30. Commonwealth Agricultural Bureaux, Edinburgh.

SORENSEN D.A., KENNEDY B.W., 1984. Estimation of response to selection using least squares and mixed model methodology. *J. Anim. Sci.*, **58**, 1097-1106.

THOMPSON R., 1977. The estimation of heritability with unbalanced data. II. Data available on more than two generations. *Biometrics*, **33**, 497-504.

THOMPSON R., 1979. Sire Evaluation. *Biometrics*, **35**, 339-353.

Appendix

In this appendix estimators of year effects and mean genetic merit are derived for the two designs.

Design I

Mixed model equations (HENDERSON, 1973) for this design have a simple form because of the pattern in A^{-1}. Let $G = \sigma_e^2/\sigma_A^2 = (1 - h^2)/h^2$, $H = G/3$ and $F = 1 + 4H = (4 - h^2)/3h^2 = 1/h_w^2$, then estimators of m_i, s_{1i} and S_{tij} satisfy

$$snm_1 + sn\bar{s}_1 = sn\bar{y}_1$$

$$snm_t + sn\bar{s}_t = sn\bar{y}_t \qquad (t = 2, \ldots, T)$$

$$m_1 + (1 + G + nH)s_{1i} - 2nH\, s_{2i} = y_{1i} \qquad (i = 1, \ldots, s)$$

$$m_1 + (1 + G)s_{1i} = y_{1i} \qquad (i = s + 1, \ldots, sn)$$

$$m_t - 2Hs_{(t-1)i1} + (F + nH)s_{ti1} - 2nHs_{(t+1)i1} = y_{ti1} \qquad (t = 2, \ldots, T - 1)$$

$$m_t - 2Hs_{(t-1)i1} + Fs_{tij} = y_{tij} \qquad (t = 2, \ldots, T - 1), (j = 2, \ldots, n)$$

$$m_t - 2Hs_{(T-1)i1} + Fs_{Tij} = y_{Tij}$$

These equations can be thought of as least squares equations with extra coefficients, (i) for males with no sons (G or F − 1 depending on whether their fathers are measured), (ii) for males with sons (G + nH or F − 1 + nH depending on whether fathers are measured), (iii) for sires and sons (− 2H).

By adding together equations and dividing by sn it can be shown that

$$m_1 + (1 + G)\bar{s}_1 + H\bar{s}_1^* - 2H\bar{s}_2 = \bar{y}_1$$

$$m_t - 2H\bar{s}_{t-1}^* + F\bar{s}_t + H\bar{s}_t^* - 2H\bar{s}_{t-1} = \bar{y}_t$$

$$m_T - 2H\bar{s}_{T-1}^* + F\bar{s}_T = \bar{y}_T$$

Hense

$$(F - 1)\bar{s}_T = 2H\bar{s}_{T-1}^* \text{ or } 2\bar{s}_T = \bar{s}_{T-1}^*$$

$$(F - 1)\bar{s}_t - 2H\bar{s}_{t-1}^* = H(\bar{s}_t^* - 2\bar{s}_{t+1})$$

$$G\,\bar{s}_1 = H(\bar{s}_1^* - 2\bar{s}_2)$$

$$\text{or } 2\bar{s}_t = \bar{s}_{t-1}^* \qquad (t = 2, \ldots, T)$$

$$\bar{s}_1 = 0$$

By adding together equations for animals not selected it can be shown that

$$s\,(n-1)m_1 + (1+G)\,(sn\bar{s}_1 - s\bar{s}_1^*) = sn\bar{y}_1 - s\bar{y}_1^*$$
$$s\,(n-1)m_t - 2Hs(n-1)\bar{s}_{t-1} + F(sn\bar{s}_t - s\bar{s}_t^*) = sn\bar{y}_t - s\bar{y}_t^*$$

or

$$(n-1)\,(\bar{y}_1 - \bar{s}_1) + (1+G)\,(n\bar{s}_1 - \bar{s}_1^*) = n\bar{y}_1 - \bar{y}_1^*$$
$$(n-1)\,(\bar{y}_t - \bar{s}_t) - 2H(n-1)\,\bar{s}_{t-1}^* + F(n\bar{s}_t - \bar{s}_t^*) = n\bar{y}_t - \bar{y}_t^*$$

and

$$(1+G)\bar{s}_1^* = (\bar{y}_1^* - \bar{y}_1)$$
$$F\,\bar{s}_t^* = (\bar{y}_t^* - \bar{y}_t) + F\,\bar{s}_t$$

Hence

$$\bar{s}_1^* = h^2\,(\bar{y}_1^* - \bar{y}_1)$$

and

$$\bar{s}_t^* = h_w^2\,(\bar{y}_t^* - \bar{y}_t) + \bar{s}_t$$

so that the mean merit of animals in year t is half the mean merit of their fathers. The merit of selected sires in year t is the mean merit in year t plus the selection differential times a measure of heritability.

Design II

The mixed model equations are now

$$snm_1 + sn\bar{s}_1 = sn\bar{y}_1$$
$$snm_2 + sn\bar{s}_2 = sn\bar{y}_2$$
$$2snm_3 + sn(\bar{s}_{13} + \bar{s}_3) = sn(\bar{y}_{13} + \bar{y}_2)$$
$$m_1 + (1 + G + 2nH)s_{1i} - 2nH\bar{s}_{2i} - 2nH\bar{s}_{13i} = y_{1j} \qquad (i = 1, \ldots, s)$$
$$m_1 + (1+G)s_{1i} = y_{1i} \qquad (i = s+1, \ldots, sn)$$
$$m_2 - 2Hs_{1i} + (F + nH)s_{2i1} - 2nH\bar{s}_{3i} = y_{2i1}$$
$$m_2 - 2Hs_{1i} + Fs_{2ij} = y_{2ij} \qquad (J = 2, \ldots, n)$$
$$m_3 - 2Hs_{1i} + Fs_{13ij} = y_{13ij}$$
$$m_3 - 2Hs_{2il} + Fs_{3ij} = y_{3ij}$$

Eliminating effects for males with no progeny and adding within generations it can be shown that

$$(1 + nG)m_1 + (1+G)\bar{s}_1^* = nG\bar{y}_1 + \bar{y}_1^* \qquad \text{(A1)}$$

$$(1 + 4nG)m_2 + 2\,(n-1)H\bar{s}_1^* + F\bar{s}_2^* = 4nH\bar{y}_2 + \bar{y}_2^* \qquad \text{(A2)}$$

$$4m_3 + (\bar{s}_1^* + \bar{s}_2^*) = 2(\bar{y}_{13} + \bar{y}_3) \qquad \text{(A3)}$$

$$m_2 + (2Hn/F)\,m_3 - 2H\bar{s}_1^* + (F + nH/F)\,\bar{s}_2^* = \bar{y}_2^* + (2Hn/F)\,\bar{y}_3 \qquad \text{(A4)}$$

$$m_1 + 2H(n-1)/F)m_2 + (2Hn/F)m_3 + [F + (2n-1)H/F]\bar{s}_1^* - 2H\bar{s}_2^* = y_1^* + (2Hn/F)\bar{y}_2 - (2H/F)\bar{y}_2^* + (2Hn/F)\bar{y}_{13} \qquad \text{(A5)}$$

Manipulating these equations (A1-5) it can be shown that

$$m_1 = \bar{y}_1 - q_1 z(h^2)$$

$$m_2 = \bar{y}_2 - (1/2)\, h^2\, (\bar{y}^*_1 - \bar{y}_1) - q_2 z(h^2)$$

$$m_3 = (\bar{y}_{13} + \bar{y}_3)/2 - (3/8)\, h^2\, (\bar{y}^*_1 - \bar{y}_1) - (1/4)\, h^2_w\, (\bar{y}^*_2 - \bar{y}_2) - q_3 z(h^2)$$

$$\bar{s}^*_1 = h^2(\bar{y}^*_1 - \bar{y}_1) - q_4 z\,(h^2)$$

$$\bar{s}^*_2 = (1/2)\, h^2\, (\bar{y}^*_1 - \bar{y}_1) + h^2_w\, (y^*_2 - \bar{y}_2) - q_5 z(h_2)$$

where $z\,(h^2)$ is defined in equation (1) and q_1, q_2, q_3, q_4 and q_5 are solutions for m_1, m_2, m_3, $\bar{s}^*_1$ and $\bar{s}^*_2$ in equation (A1) – (A5) with $\bar{y}_1 = \bar{y}^*_1 = \bar{y}_2 = \bar{y}^*_2 = 0$ and $\bar{y}_{13} = \bar{y}_3 = -1/2$.

Hence the q values are functions of h^2 and n. The mean genetic values can be derived from the estimates of m_1, m_2 and m_3 and are given in equation (2) – (4).

27

The maximum likelihood approach to the estimate of liability

by

R. Thompson

Ann. Hum. Genet., Lond. (1972), **36**, 221
Printed in Great Britain

The maximum likelihood approach to the estimate of liability

By R. THOMPSON

ARC Unit of Statistics, University of Edinburgh, 21 Buccleuch Place, Edinburgh EH 8 9LN, Scotland

INTRODUCTION

Falconer (1965) has applied methods developed in quantitative genetics for dealing with threshold characters to the problem of the inheritance of liabilities to diseases. In particular he showed how data on the prevalence of a disease in the general population and among relatives of affected individuals can be used to measure the correlation between relatives for a disease. Falconer surmounts the difficulty of the all-or-none character of a disease by assuming an underlying continuous liability to a disease. This liability is supposed to consist of many small genetic and environmental effects and so is approximated by a normal distribution. There is a critical point, or threshold, on the liability scale above which individuals are affected and below which individuals are normal. Provided that the non-genetic causes of resemblance can be removed from the calculated correlation between relatives, an estimate of the heritability of liability to a disease can be found.

Edwards (1969) pointed out that Falconer's estimates of correlation are biased downwards by up to 10 %. He also gives an alternative method of estimating the correlation by using tetrachoric functions for bivariate normal functions.

Falconer's and Edwards's methods are analogous to estimating heritabilities in quantitative trials from the response to selection, and from parent–offspring regressions, respectively. We will instead investigate the maximum-likelihood method approach. We first consider the case when data are available on equally related groups of size n. If we again take the analogy with quantitative genetics, this corresponds to the case when sib correlation techniques are used to estimate heritabilities. To adapt Falconer's and Edwards's methods for this estimation would not make use of all available information. We develop the relevant estimating equations and compare the sampling variance of the correlation between relatives for the maximum-likelihood method with those for Edwards's method for one particular situation.

We then show how the maximum-likelihood procedure can be modified to deal with situations similar to those discussed by Falconer. Falconer considers a more general case that can be thought of as follows: we observe the prevalence of a disease in N unrelated individuals (propositi) and then observe the prevalence in relatives of N_1 of these normal and N_2 of these affected individuals.

The special case when we observe the relatives of all N propositi (i.e. $N_1+N_2=N$) is equivalent to the situation mentioned earlier; that is, when data is available on unselected families of size n.

Since the variance of the correlation between relatives depends on n, N_1, N_2 and N it is natural to ask what experimental designs are most efficient for the estimation of the correlation between relatives. Robertson (1959), with sib correlation methods, and Latter & Robertson (1960), with parent–offspring techniques, have shown the advantage of choosing the family size optimally when estimating heritabilities in quantitative traits.

Making use of the fact that the variance of the estimate of correlation depends on N_1 and N_2

is analogous with a suggestion of Hill (1970). He notes that the precision of heritabilities from parent–offspring regressions can be improved if extreme individuals are mated.

For laboratory or farm animals it might be feasible to arrange the family structure in some optimum manner but obviously there is little scope for this with human populations.

Since no analytic solution could be found, optimum designs over a practical range of thresholds, family sizes and correlations have been identified numerically and these indicate the usefulness of the procedure.

MAXIMUM-LIKELIHOOD APPROACH

We first consider the case when data are available from N groups of size n. We let $U_1, U_2, \ldots, U_n$ be n standardized variates representing the underlying variates for the 1st, 2nd, ..., nth individuals in a group. We assume $U_1, U_2, \ldots, U_n$ have a joint multivariate normal density function

$$\phi(u_1, u_2, \ldots, u_n; \rho) = \phi(\mathbf{u}; \rho) = (2\pi)^{-n/2} |\mathbf{V}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{u}'\mathbf{V}^{-1}\mathbf{u})},$$

where $\mathbf{u}$ is a $n \times 1$ vector with ith element u_i $(i = 1, \ldots, n)$ and $\mathbf{V}$ is a $n \times n$ matrix with diagonal elements 1 and off-diagonal elements ρ. We assume that the correlation between pairs of underlying liabilities is the same and is represented by ρ.

We define n variates $X_1, X_2, \ldots, X_n$ such that $X_i = 0$ if $U_i < a_i$ and $X_i = 1$ if $U_i \geqslant a_i$ $(i = 1, \ldots, n)$. So $X_i = 1$ if the ith individual in the group is affected by the disease and $X_i = 0$ otherwise. The threshold for the underlying variate U_i is a_i. In certain cases some, or all, of the thresholds can be assumed to be equal. For example, we could possibly be considering data from groups of three individuals. The groups might then consist of a mother, one of her daughters and one of her sons and we could represent their underlying liabilities by U_1, U_2 and U_3 respectively. In some cases we might be willing to assume that $a_2 = a_3$; that is, the prevalence of the disease in children is independent of sex but might be different from the prevalence in parents. Or we might know that the prevalence is independent of age but might differ between sexes and so take $a_1 = a_2$. We discuss the case of n different thresholds but there is no difficulty in adapting the results for cases when some of the thresholds can be assumed to be equal. Then

$$\Pr(X_i = 0) = \Pr(U_i < a_i) = \int_{-\infty}^{a_i} \phi(u_i)\, du_i = \Phi(a_i)$$

and

$$\Pr(X_i = 1) = \Pr(U_i \geqslant a_i) = \int_{a_i}^{\infty} \phi(u_i)\, du_i = 1 - \Phi(a_i) \quad (i = 1, \ldots, n).$$

The joint distribution of X_i and X_j is given by

$$\Pr(X_i = 0, X_j = 0) = \int_{-\infty}^{a_i} \int_{-\infty}^{a_j} \phi(u_i, u_j; \rho)\, du_i\, du_j = \Phi(a_i, a_j; \rho),$$

$$\Pr(X_i = 0, X_j = 1) = \Phi(a_i) - \Phi(a_i, a_j; \rho),$$

$$\Pr(X_i = 1, X_j = 0) = \Phi(a_j) - \Phi(a_i, a_j; \rho),$$

$$\Pr(X_i = 1, X_j = 1) = 1 - \Phi(a_i) - \Phi(a_j) + \Phi(a_i, a_j; \rho) \quad (i, j = 1, \ldots, n).$$

We can conveniently write the joint distribution of $X_1, X_2, \ldots, X_n$ by defining

$$P_{\mathbf{k}} = \Pr(X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n),$$

where $\mathbf{k}$ is a $n \times 1$ vector with ith element k_i.

Since k_i can only take the values 0 or 1, there are only 2^n possible values for $\mathbf{k}$ and we let this set of vectors be denoted by K.

Then $P_k = \Phi_1(\mathbf{a}, \mathbf{k}; \rho)$ $(\mathbf{k} \in K)$ where the function $\Phi_1(\mathbf{a}, \mathbf{k}; \rho)$ means that u_i is integrated from $-\infty$ to a_i if the ith element of $\mathbf{k}$ is 0 and from a_i to $+\infty$ if the ith element of $\mathbf{k}$ is 1 $(i = 1, \ldots, n)$.

We define a similar function $\Phi_2(\mathbf{a}, \mathbf{k}; \rho)$, which denotes that u_i is integrated from $-\infty$ to a_i if the ith element of $\mathbf{k}$ is 0 and from a_i to $+\infty$ if the ith element of $\mathbf{k}$ is 1 $(i = 1, \ldots, n)$.

$\Phi_2(\mathbf{a}, \mathbf{k}; \rho)$ then reduces to a $n - b(\mathbf{k})$ dimensional integral, where $b(\mathbf{k})$ is the number of elements that are 1 in the vector $\mathbf{k}$. $b(\mathbf{k})$ can be evaluated as $b(\mathbf{k}) = \mathbf{1}'\mathbf{k}$ where $\mathbf{1}'$ is a $1 \times n$ vector of ones.

In a similar manner we let

$$\phi_2(\mathbf{a}, \mathbf{m}; \rho) = \phi(a_i), \quad \text{where} \quad \mathbf{m} = \mathbf{I}(i)$$

and

$$\phi_2(\mathbf{a}, \mathbf{m}; \rho) = \phi(a_i, a_j; \rho), \quad \text{where} \quad \mathbf{m} = \mathbf{I}(i) + \mathbf{I}(j) \quad (i, j = 1, \ldots, n, i \neq j),$$

where $\mathbf{I}(i)$ is a $n \times 1$ vector with all elements zero except the ith element, which is 1.

Since $\Phi_2(\mathbf{a}, \mathbf{k}; \rho)$ reduces to a $(n - b(\mathbf{k}))$ dimensional integral we find it more convenient to express the probabilities $P_\mathbf{k}$ in terms of the Φ_2 integrals rather than the Φ_1 integrals.

It is easy to see that when $n = 2$

$$\Phi_2(\mathbf{a}, (0,0); \rho) = \Phi_1(\mathbf{a}, (0,0); \rho),$$

$$\Phi_2(\mathbf{a}, (0,1); \rho) = \Phi_1(\mathbf{a}, (0,0); \rho) + \Phi_1(\mathbf{a}, (0,1); \rho),$$

$$\Phi_2(\mathbf{a}, (1,0); \rho) = \Phi_1(\mathbf{a}, (0,0); \rho) + \Phi_1(\mathbf{a}, (1,0); \rho),$$

$$\Phi_2(\mathbf{a}, (1,1); \rho) = \Phi_1(\mathbf{a}, (0,0); \rho) + \Phi_1(\mathbf{a}, (0,1); \rho) + \Phi_1(\mathbf{a}, (1,0); \rho) + \Phi_1(\mathbf{a}, (1,1); \rho).$$

This can be generalized to

$$\Phi_2(\mathbf{a}, \mathbf{k}; \rho) = \sum_{\mathbf{j} \in K(\mathbf{k})} \Phi_1(\mathbf{a}, \mathbf{j}; \rho) \quad (\mathbf{k} \in K),$$

where $K(\mathbf{k})$ denotes the set of vectors $\mathbf{j}$ (contained in K) such that the ith element of $\mathbf{j}$ is less than or equal to the ith element of $\mathbf{k}$.

These give 2^n simultaneous equations for the Φ_1 integrals in terms of the Φ_2 integral. We can reorganize the equations so that

$$\Phi_1(\mathbf{a}, \mathbf{B}(j); \rho) = \Phi_2[\mathbf{a}, \mathbf{B}(j); \rho] - \sum_{i=0}^{j-1} \Phi_1[\mathbf{a}, \mathbf{B}(i); \rho] A'_{ij} \quad (i = 0, \ldots, 2^n - 1),$$

where the $n \times 1$ vector $\mathbf{B}(j)$ has elements 0 or 1 (and so included in K) and the elements b_i are specified by the equation

$$\mathbf{j} = \sum_{i=1}^{n} b_i \times 2^{i-1}, \quad \text{and} \quad A'_{ij} = A_{\mathbf{B}(i)\mathbf{B}(j)},$$

where $A_{\mathbf{B}(i)\mathbf{B}(j)} = 1$ if the term-by-term or Schur-Hadamard product $\mathbf{B}(i).\mathbf{B}(j) = \mathbf{B}(i)$ and $A_{\mathbf{B}(i)\mathbf{B}(j)} = 0$ if $\mathbf{B}(i).\mathbf{B}(j) \neq \mathbf{B}(i)$.

So we can simply calculate the Φ_1 integrals related to $\mathbf{B}(0)$, $\mathbf{B}(1)$, ..., $\mathbf{B}(2^n - 1)$ in turn. When $n = 3$ the integrals can be calculated in the order (0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 1, 0), (0, 0, 1), (1, 0, 1), (0, 1, 1), (1, 1, 1).

If the equations are being handled for a computer it is sometimes convenient to note that the single number j expressed in binary form (i.e. expressed in powers of 2) can be used to represent the vector $\mathbf{B}(j)$. In some computer languages, Fortran for example, the term-by-term product can be simulated using the AND function (Louden, 1967, p. 246), where

$$\mathbf{B}[\text{AND}(j, k)] = \mathbf{B}(j).\mathbf{B}(k).$$

The relationship between the Φ_1 and Φ_2 integrals is analogous to the one between the crude sum of squares and corrected sum of squares in an analysis of variance. Part of an algorithm of Gower (1969) could therefore be used to transform from the $\Phi_2(\mathbf{a}, \mathbf{k}; \rho)$ to the $\Phi_1(\mathbf{a}, \mathbf{k}; \rho)$. Gower notes that such a transformation is similar to that of finding treatment effects in a 2^n experiment from individual values by Yates's algorithm. Gower (1969) uses scanning routines for simulating multidimensional arrays (Gower, 1968) to implement this, but the transformation can easily be programmed without recourse to these routines.

We want to estimate $\mathbf{a}$ and ρ from the Nn observations and we first consider the N groups according to their values of $X_1, X_2, \ldots, X_n$. There are 2^n different possibilities and, just as we have defined $P_{\mathbf{k}}$, we can define $n_{\mathbf{k}}$ to be the number of groups such that X_i is k_i $(i = 1, \ldots, n; \mathbf{k} \in K)$.

The logarithmic likelihood can be written as

$$\text{const} + \sum_{\mathbf{k} \in K} n_{\mathbf{k}} \ln P_{\mathbf{k}}.$$

The maximum-likelihood estimates of a_i and ρ satisfy

$$\sum_{\mathbf{k} \in K} \frac{n_{\mathbf{k}}}{P_{\mathbf{k}}} \frac{\partial P_{\mathbf{k}}}{\partial a_i} = 0 \quad (i = 1, \ldots, n) \tag{1}$$

and

$$\sum_{\mathbf{k} \in K} \frac{n_{\mathbf{k}}}{P_{\mathbf{k}}} \frac{\partial P_{\mathbf{k}}}{\partial \rho} = 0. \tag{2}$$

Since we are considering a multinomial distribution we can easily write down the information matrix (for instance Bailey (1961), p. 277). We find that the symmetric information matrix of a_i $(i = 1, \ldots, n)$ and ρ has elements $I_{s,t}$ where

$$I_{s,t} = N \sum_{\mathbf{k} \in K} \frac{1}{P_{\mathbf{k}}} \frac{\partial P_{\mathbf{k}}}{\partial a_s} \frac{\partial P_{\mathbf{k}}}{\partial a_t} \quad (1 \leqslant s, t \leqslant n),$$

$$I_{s,n+1} = N \sum_{\mathbf{k} \in K} \frac{1}{P_{\mathbf{k}}} \frac{\partial P_{\mathbf{k}}}{\partial a_s} \frac{\partial P_{\mathbf{k}}}{\partial \rho} \quad (1 \leqslant s \leqslant n), \tag{3}$$

and

$$I_{n+1,n+1} = N \sum_{\mathbf{k} \in K} \frac{1}{P_{\mathbf{k}}} \left(\frac{\partial P_{\mathbf{k}}}{\partial \rho} \right)^2.$$

In the appendix we show how the differentials of $P_{\mathbf{k}}$ with respect to a_i and ρ can be expressed in terms of lower dimensional multinormal integrals.

The usual iterative approach to the solution of (1) and (2) has been found quite satisfactory in practice. Good initial estimates can be found for the threshold values by converting the mean values of the X variates to standard normal deviates.

A SIMPLER APPROACH

For the case when there are only two thresholds involved, say a_1 associated with the first individual of a group and a_2 connected with its relatives, a simpler approach and similar to that of Edwards and Falconer is possible. We might classify the number of unaffected and affected relatives according to whether or not the first individual in the group is affected, i.e. by the two-way table

	$X_1 = 0$	$X_1 = 1$
Number of relatives unaffected...	M_0	M_1
Number of relatives affected ...	M_2	M_3

where $M_0 = \sum_{\mathbf{k} \in K_0} n_{\mathbf{k}}(n - 1 - b(\mathbf{k}))$, $M_1 = \sum_{\mathbf{k} \in K_1} n_{\mathbf{k}}(n - 2 - b(\mathbf{k}))$,

and $$M_2 = \sum_{\mathbf{k}\in K_0} n_{\mathbf{k}} b(\mathbf{k}), \quad M_3 = \sum_{\mathbf{k}\in K_1} n_{\mathbf{k}}(b(\mathbf{k})-1),$$

where K_0 and K_1 are the sets of vectors $\mathbf{k}\in K$ such that $\mathbf{k}\in K_0$ if $k_1 = 0$ and $\mathbf{k}\in K_1$ if $k_1 = 1$.

In this formulation there are three parameters to be estimated based on the four elements of the two-way table. It is natural to equate observed and expected frequencies to give estimating equations for a_1, a_2 and ρ. This, in particular, is the maximum-likelihood procedure if all other information except the two-way table has been lost or censored.

These reduce to

$$\Phi(a_1) = \frac{M_0+M_2}{Nn}, \tag{4}$$

$$\Phi(a_2) = \frac{M_0+M_1}{Nn}, \tag{5}$$

and $$\Phi(a_1, a_2; \rho) = \frac{M_0}{Nn}. \tag{6}$$

Since this method of estimation is essentially equivalent to that suggested by Edwards it seems useful to compare the precision of the correlation estimate from this method with those from the maximum-likelihood method.

By first finding the sampling variances of N_0, N_1 and N_2, the variances and covariances of the estimates of a_1, a_2 and ρ from (4)–(6) can be found to satisfy

$$N\left[\frac{\partial\Phi(a_1)}{\partial a_1}\right]^2 \operatorname{var}(a_1) = \Phi(a_1) - [\Phi(a_1)]^2,$$

$$N\frac{\partial\Phi(a_1)}{\partial a_1}\frac{\partial\Phi(a_2)}{\partial a_2}\operatorname{cov}(a_1, a_2) = B - \Phi(a_1)\,\Phi(a_2),$$

$$N\left[\frac{\partial\Phi(a_2)}{\partial a_2}\right]^2 \operatorname{var}(a_2) = (n-1)[\Phi(a_2) + (n-2)C - (n-1)[\Phi(a_2)]^2],$$

$$N\frac{\partial\Phi(a_1)}{\partial a_1}\frac{\partial B}{\partial\rho}\operatorname{cov}(a_1,\rho) + \frac{\partial B}{\partial a_1}\operatorname{var}(a_1) + \frac{\partial B}{\partial a_2}\operatorname{cov}(a_1, a_2) = B[1-\Phi(a_1)],$$

$$\frac{\partial\Phi(a_2)}{\partial a_2}\left[\frac{\partial B}{\partial\rho}\operatorname{cov}(a_2,\rho) + \frac{\partial B}{\partial a_1}\operatorname{cov}(a_1,\rho) + \frac{\partial B}{\partial a_2}\operatorname{var}(a_2)\right] = \frac{A-(n-1)^2 B\Phi(a_2)}{N(n-1)^2}$$

and

$$\left[\frac{\partial B}{\partial\rho}\right]^2 \operatorname{var}(\rho) + 2\frac{\partial B}{\partial\rho}\frac{\partial B}{\partial a_1}\operatorname{cov}(a_1,\rho) + \left[\frac{\partial B}{\partial a_1}\right]^2 \operatorname{var}(a_1) + 2\frac{\partial B}{\partial a_1}\frac{\partial B}{\partial a_2}\operatorname{cov}(a_1, a_2) + 2\frac{\partial B}{\partial\rho}\frac{\partial B}{\partial a_2}\operatorname{cov}(a_2,\rho)$$
$$+\left[\frac{\partial B}{\partial a_2}\right]^2 \operatorname{var}(a_2) = \frac{A-(n-1)^2 C^2}{N(n-1)^2},$$

where $$A = \sum_{\mathbf{k}\in K_0} [n-1-b(\mathbf{k})]^2 P_{\mathbf{k}},$$

$$B = \Phi(a_1, a_2; \rho) \quad \text{and} \quad C = \Phi(a_2, a_2; \rho).$$

The information matrix for the maximum-likelihood estimates of a_1, a_2 and ρ can be derived from the terms defined in (3) and has elements $I^*_{s,t}$, where

$$I^*_{1,1} = I_{1,1},$$

$$I^*_{1,2} = \sum_{s=2}^{n} I_{1,s}, \quad I^*_{1,3} = I_{1,n+1},$$

$$I^*_{2,2} = \sum_{s=2}^{n}\sum_{t=2}^{n} I_{s,t}, \quad I^*_{2,3} = \sum_{s=2}^{n} I_{2,n+1},$$

$$I^*_{3,3} = I_{3,3}.$$

Table 1. *Variances* ($\times 10^5$) *of estimates of the correlation coefficient for* 1000 *individuals, estimated by maximum-likelihood (on first line) and by using equations* (4)–(6) *(on second line), for various values of the correlation* ($\rho = 0{\cdot}1\,(0{\cdot}2)\,0{\cdot}5$), *of the number in the group* ($n = 2\,(2)\,6$) *and of the threshold* ($a = 1{\cdot}0\,(0{\cdot}4)\,3{\cdot}0$)

	$\rho = 0{\cdot}1$			$\rho = 0{\cdot}3$			$\rho = 0{\cdot}5$		
a	$n = 2$	$n = 4$	$n = 6$	$n = 2$	$n = 4$	$n = 6$	$n = 2$	$n = 4$	$n = 6$
1·0	982	388	272	818	431	372	594	393	386
		716	698		702	773		584	695
1·4	1,967	747	506	1,520	756	626	1,048	663	631
		1,405	1,349		1,270	1,375		1,013	1,194
1·8	5,097	1,861	1,217	3,503	1,628	1,285	2,237	1,334	1,223
		3,567	3,363		2,832	2,995		2,108	2,453
2·2	17,388	6,136	3,892	10,103	4,378	3,278	5,792	3,231	2,839
		11,951	11,080		7,865	8,096		5,292	6,049
2·6	79,220	27,258	16,861	36,806	14,917	10,609	18,274	9,483	7,952
		53,711	49,169		27,588	27,579		16,086	17,992
3·0	486,710	165,114	100,880	170,812	65,286	44,240	70,562	34,039	27,202
		326,826	297,051		123,666	120,128		59,666	65,071

The variances of the estimates of the correlation coefficient for the two methods have been evaluated for the case when $a_1 = a_2 = a$ and the total number of individuals measured (Nn) is 1000. These are tabulated in Table 1 for various values of the number in the group, the correlation and the threshold groups; sizes from 2 to 6 are used since, at least for the first-degree relatives (for example full sibs), most family groups are in this range. Correlations from 0·1 to 0·5 are tabled. In the absence of any environmental correlation a correlation of 0·5 for between first degree relatives represents a heritability of 1. Thresholds range from 1·0 to 3·0 representing prevalences in the whole population of 15·9 % to 0·13 %.

When $n = 2$ of course the two methods are identical. Table 1 shows the benefit of using the maximum-likelihood approach especially when $n \geqslant 4$. For small correlations, over the range of group sizes and thresholds considered, the efficiency of using equations (4)–(6) compared with the maximum-likelihood is roughly $2/n$, the approximation being relatively better for larger thresholds.

For the ranges of thresholds and correlations considered the optimum group size, given that a fixed number of individuals are measured, to minimize the variance of the maximum-likelihood estimate of the correlation coefficient is equal to or greater than 6 (except when $a = 1$ and $\rho = 0{\cdot}5$ when the optimum is $n = 5$). This conclusion is similar to results of Robertson (1959) in the quantitative situation, but there seems little practical use in evaluating the optimum n in our situation.

FALCONER'S CASE

Falconer considers a more complicated situation that corresponds to the case when we have observations on N unrelated individuals (propositi), say, associated with the a_1 thresholds, but observes relatives of only N_1 of the N' normal and of N_2 of the N'' affected propositi ($N' + N'' = N$). One common case is when we observe relatives only of infected individuals, i.e. $N_1 = 0$, $N_2 = N''$. We will suppose we take observations on $(n-1)$ relatives so that the group size is again n. The log likelihood in this situation can be written, using the same notation as before, as

$$\text{const} + \sum_{\mathbf{k} \in K} n_{\mathbf{k}} \ln P_{\mathbf{k}} + (N' - N_1) \ln [\Phi(a_1)] + (N'' - N_2) \ln (1 - \Phi(a_1)).$$

Table 2. *Optimum proportion* $(N_1+N_2)/N$ *and variances* ($\times 10^5$) *of estimates of the correlation from* 1000 *individuals, for various values of the correlation* ($\rho = 0{\cdot}1\ (0{\cdot}2)\ 0{\cdot}5$), *of the number in the group* ($n = 2\ (2)\ 6$) *and of the threshold* ($a = 1{\cdot}0\ (0{\cdot}4)\ 3{\cdot}0$) *for case* (*i*)

	$\rho = 0{\cdot}1$			$\rho = 0{\cdot}3$			$\rho = 0{\cdot}5$		
a	$n = 2$	$n = 4$	$n = 6$	$n = 2$	$n = 4$	$n = 6$	$n = 2$	$n = 4$	$n = 6$
1·0	0·617	0·561	0·525	0·675	0·558	0·500	0·737	0·568	0·499
	912	364	256	778	398	337	576	360	341
1·4	0·409	0·356	0·325	0·476	0·374	0·323	0·548	0·396	0·332
	1,630	638	438	1,323	634	512	953	551	497
1·8	0·262	0·216	0·192	0·331	0·249	0·208	0·408	0·280	0·224
	3,726	1,421	949	2,754	1,237	948	1,880	1,010	864
2·2	0·163	0·128	0·111	0·229	0·167	0·136	0·307	0·204	0·158
	11,339	4,187	2,722	7,191	3,041	2,218	4,494	2,257	1,840
2·6	0·098	0·074	0·063	0·158	0·113	0·091	0·233	0·153	0·117
	47,110	16,854	10,668	23,982	9,605	6,705	13,142	6,220	4,863
3·0	0·057	0·042	0·035	0·107	0·076	0·061	0·177	0·117	0·089
	270,537	94,568	59,017	103,354	39,458	26,454	47,359	21,238	15,996

Table 3. *Optimum proportion* $(N_1+N_2)/N$ *and variances* ($\times 10^5$) *of estimates of the correlation from* 1000 *individuals for various values of the correlation* ($\rho = 0{\cdot}1\ (0{\cdot}2)\ 0{\cdot}5$), *of the number in the group* ($n = 2\ (2)\ 6$) *and of the threshold* ($a = 1{\cdot}0\ (0{\cdot}4)\ 3{\cdot}0$) *for the case* (*ii*)

	$\rho = 0{\cdot}1$			$\rho = 0{\cdot}3$			$\rho = 0{\cdot}5$		
a	$n = 2$	$n = 4$	$n = 6$	$n = 2$	$n = 4$	$n = 6$	$n = 2$	$n = 4$	$n = 6$
1·0	0·159	0·361	0·388	0·200	0·305	0·308	0·282	0·248	0·237
	807	349	249	709	372	318	539	327	308
1·4	0·080	0·080	0·125	0·080	0·080	0·080	0·080	0·080	0·080
	1,289	549	397	1,061	524	433	784	438	390
1·8*	2,906	1,121	767	2,114	944	725	1,449	744	622
2·2*	9,201	3,304	2,130	5,580	2,288	1,645	3,407	1,630	1,292
2·6*	40,483	13,950	8,638	19,337	7,428	5,064	10,094	4,550	3,458
3·0*	245,174	83,072	50,776	87,409	31,944	20,891	37,450	16,018	11,739

* The optimum proportion $(N_1+N_2)/N$ for thresholds 1·8, 2·2, 2·6 and 3·0 are 0·0359, 0·0139, 0·0047 and 0·0013 respectively.

The estimating equations become

$$\sum_{\mathbf{k}\in K}\frac{n_{\mathbf{k}}}{P_{\mathbf{k}}}\frac{\partial P_{\mathbf{k}}}{\partial a_i}+\delta_{i1}\left(\frac{N'-N_1}{\Phi(a_1)}-\frac{N''-N_2}{1-\Phi(a_1)}\right)\frac{\partial\Phi(a_1)}{\partial a_1}=0 \quad (i=1,\ldots,n), \tag{7}$$

where $\delta_{i1}=1$ if $i=1$ and zero otherwise and

$$\sum_{\mathbf{k}\in K}\frac{n_{\mathbf{k}}}{P_{\mathbf{k}}}\frac{\partial P_{\mathbf{k}}}{\partial\rho}=0.$$

The symmetric information matrix of a_i $(i=1,\ldots,n)$ and ρ has elements $I_{s,t}$, where

$$I_{s,t}=N\left[\sum_{\mathbf{k}\in K_0}\left(\frac{P'}{P_{\mathbf{k}}}\frac{\partial P_{\mathbf{k}}}{\partial a_s}\frac{\partial P_{\mathbf{k}}}{\partial a_t}\right)+\sum_{\mathbf{k}\in K_1}\left(\frac{P''}{P_{\mathbf{k}}}\frac{\partial P_{\mathbf{k}}}{\partial a_s}\frac{\partial P_{\mathbf{k}}}{\partial a_t}\right)\right]+\delta_{1s}\delta_{1t}\left[\frac{1-P'}{\Phi(a_1)}+\frac{1-P''}{1-\Phi(a_1)}\right]\left[\frac{\partial\Phi(a_1)}{\partial a_1}\right]^2$$

$$(s,t=1,\ldots,n),$$

$$I_{s,n+1}=N\left[\sum_{\mathbf{k}\in K_0}\left(\frac{P'}{P_{\mathbf{k}}}\frac{\partial P_{\mathbf{k}}}{\partial a_s}\frac{\partial P_{\mathbf{k}}}{\partial\rho}\right)+\sum_{\mathbf{k}\in K_1}\left(\frac{P''}{P_{\mathbf{k}}}\frac{\partial P_{\mathbf{k}}}{\partial a_s}\frac{\partial P_{\mathbf{k}}}{\partial\rho}\right)\right] \quad (s=1,\ldots,n)$$

and

$$I_{n+1,n+1}=N\left[\sum_{\mathbf{k}\in K_0}\left(\frac{P'}{P_{\mathbf{k}}}\left(\frac{\partial P_{\mathbf{k}}}{\partial\rho}\right)^2\right)+\sum_{\mathbf{k}\in K_1}\left(\frac{P''}{P_{\mathbf{k}}}\left(\frac{\partial P_{\mathbf{k}}}{\partial\rho}\right)^2\right)\right],$$

where

$$P'=N_1/N\Phi(a_1) \quad \text{and} \quad P''=N_2/N(1-\Phi(a_1)).$$

If we are able to measure a total of T individuals we might ask how do we choose N, N_1 and N_2 to minimize the variance of the estimate of the correlation between relatives. It seems more useful to treat the group size as a fixed parameter of the data along with the correlation and the thresholds rather than try to find an optimum n. In fact if we were to find optimal N, N_1, N_2 and n, then this resulting n would be larger than the optimal n for the case discussed in the previous section, i.e. when $N_1+N_2=N$. Hill (1970) found a similar result in an analogous quantitative situation.

Two situations seem worth considering (i) when $a_2=a_3=\ldots=a_n$ and we wish to estimate separately a_1 and a_2, (ii) when $a_1=a_2=\ldots=a_n$. Since no analytic solution giving the optimal values of N, N_1 and N_2 could be found the optimum proportion $(N_1+N_2)/N$ and the resulting variance of the estimate of the correlation coefficient have been found numerically, when $a=a_1=a_2$ and 1000 individuals are measured, and are given in Tables 2 and 3. The proportions and variances are tabulated for the same ranges of correlation, number in group and threshold as in Table 1. The optimal N_2 is always equal to the number of affected propositi observed (N''). Hence N can be calculated from

$$T=N[(N_1+N_2)(n-1)/N+1].$$

When we assume one common threshold (case ii) there is less need to observe relatives of normal individuals than when we estimate two thresholds (case i). In fact if the threshold is greater than 1·6, the optimum procedure in case (ii) for the group sizes considered is to just look at relatives of infected individuals, i.e. $N_1=0$.

In both cases there is a useful reduction in variance compared with the values in Table 1, the gain being greater for higher thresholds.

DISCUSSION

Edwards (1969) has suggested reasons why the threshold model might not hold. It is useful to note that with the maximum-likelihood method, provided that we are estimating less than 2^n-1 parameters, we can test the assumption of an underlying multivariate normal distribution by using a χ^2 test to compare the observed and expected frequencies.

The development has been solely in terms of a constant group size. Often, however, data from biological experiments are obtained from groups of different sizes and the question naturally arises how do we combine the information from the groups of different sizes.

Falconer (1965) and Edwards (1969) implicitly give equal weight to each relative, an optimum procedure when the correlation is zero. This and other weighting procedures have been discussed by Kempthorne & Tandon (1953) and Bohren, McKean & Yamada (1961) for quantitative traits.

However, when the group size is not constant over all groups, but is independent of the thresholds and the correlation, the maximum-likelihood approach can easily be generalized and gives a natural way of combining the information from differently sized groups.

The maximum-likelihood approach could also be extended to deal with more complex correlation structures between relatives. The main difficulty is that the evaluation of the probabilities and their differentials usually become more difficult, but this will not be discussed further here.

SUMMARY

The maximum-likelihood method is applied to the estimation of the parameters of the multifactorial model for the inheritance of disease liability (Falconer, 1965). The method is then compared with a simpler approach suggested by the work of Edwards (1969). The selection of an optimum design for estimating the correlation between relatives is discussed.

REFERENCES

Bailey, N. T. J. (1961). *Introduction to the Mathematical Theory of Genetic Linkage*. Oxford University Press.

Bohren, B. B., McKean, H. E. & Yamada, Y. (1961). Relative efficiencies of heritability estimates based on regression of offspring on parent. *Biometrics* **17**, 481–91.

Curnow, R. N. & Dunnett, C. W. (1962). The numerical evaluation of certain multivariate normal integrals. *Annals of Mathematical Statistics* **33**, 571–9.

Edwards, J. H. (1969). Familial predisposition in man. *British Medical Bulletin* **25**, 58–63.

Falconer, D. S. (1965). The inheritance of liability to certain diseases, estimated from the incidences among relatives. *Annals of Human Genetics* **29**, 51–76.

Gower, J. C. (1968). Simulating multidimensional arrays in one dimension (Algorithm AS1). *Applied Statistics* **17**, 180–5.

Gower, J. C. (1969). Analysis of variance for a factorial table (Algorithm AS19). *Applied Statistics* **18**, 199–202.

Hill, W. G. (1970). Design of experiments to estimate heritability by regression on selected parents. *Biometrics* **26**, 566–71.

Kempthorne, O. & Tandon, O. B. (1953). The estimation of heritability by regression of offspring on parent. *Biometrics* **9**, 90–100.

Latter, B. D. H. & Robertson, A. (1960). Experimental design in the estimation of heritability by regression methods. *Biometrics* **16**, 348–53.

Louden, R. K. (1967). *Programming the IBM* 1130 *and* 1800. New Jersey: Prentice-Hall Inc.

Robertson, A. (1959). Experimental design in the evaluation of genetic parameters. *Biometrics* **15**, 219–26.

Tallis, G. M. (1962). The maximum likelihood estimation of correlation from contingency tables. *Biometrics* **18**, 342–53.

APPENDIX

We show in this appendix how the differentials required in the solution of the estimating equations (1) and (2) can be obtained.

We first write P_0 in the form of a single integral involving only functions of the one-dimensional normal distribution, i.e.

$$P_0 = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \prod_{i=1}^{n} \Phi\left(\frac{a_i - \rho^{\frac{1}{2}} y}{(1-\rho)^{\frac{1}{2}}}\right) \exp\left(-\tfrac{1}{2}y^2\right) dy,$$

where

$$\Phi(t) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{t} \exp\left(-\tfrac{1}{2}y^2\right) dy.$$

This form has been suggested, for example, by Curnow & Dunnett (1962). Then

$$\frac{\partial P_0}{\partial \rho} = \frac{1}{2} \sum_{l=1}^{n} (2\pi)^{-1} \int_{-\infty}^{\infty} \left[\prod_{\substack{i=1 \\ i \neq l}}^{n} \Phi\left(\frac{a_i - \rho^{\frac{1}{2}} y}{(1-\rho)^{\frac{1}{2}}}\right)\right] (a_l - \rho^{-\frac{1}{2}} y)\, (1-\rho)^{-\frac{3}{2}} \exp\left[\frac{-\frac{1}{2}(a_l - \rho^{\frac{1}{2}} y)^2}{1-\rho} - \tfrac{1}{2}y^2\right] dy$$

$$= \frac{1}{2} \sum_{l=1}^{n} (2\pi)^{-1} \exp\left(-\tfrac{1}{2}a_l^2\right) \int_{-\infty}^{\infty} \left[\prod_{\substack{i=1 \\ i \neq l}}^{n} \Phi\left(\frac{a_i - \rho^{\frac{1}{2}} y}{(1-\rho)^{\frac{1}{2}}}\right)\right] (a_l - \rho^{-\frac{1}{2}} y)\, (1-\rho)^{\frac{3}{2}} \exp\left[\frac{-\frac{1}{2}(y - \rho^{\frac{1}{2}} a_l)^2}{1-\rho}\right] dy.$$

Integrating by parts we can write $\partial P_0 / \partial \rho$ as

$$= \frac{1}{2} \sum_{l=1}^{n} (2\pi)^{-\frac{3}{2}} \exp\left(-\tfrac{1}{2}a_l^2\right) \sum_{\substack{j=1 \\ j \neq l}}^{n} \int_{-\infty}^{\infty} \left[\prod_{\substack{i=1 \\ i \neq j, l}}^{n} \Phi\left(\frac{a_i - \rho^{\frac{1}{2}} y}{(1-\rho)^{\frac{1}{2}}}\right)\right] (1-\rho)^{-1} \exp\left[\frac{-\frac{1}{2}[(a_j - \rho^{\frac{1}{2}} y)^2 + (y - \rho^{\frac{1}{2}} a_l)^2]}{1-\rho}\right] dy$$

$$= \frac{1}{2} \sum_{l=1}^{n} \sum_{\substack{j=1 \\ j \neq l}}^{n} (2\pi)^{-1} (1-\rho^2)^{-\frac{1}{2}} \exp\left[\frac{-\frac{1}{2}[a_l^2 + a_j^2 - 2\rho a_l a_j]}{1-\rho^2}\right]$$

$$\times \left[\int_{-\infty}^{\infty} \left[\prod_{\substack{i=1 \\ i \neq j, l}}^{n} \Phi\left(\frac{a_i - \rho^{\frac{1}{2}} y}{(1-\rho)^{\frac{1}{2}}}\right)\right] \exp\left[\frac{-\frac{1}{2}(1+\rho)\,[y - \rho^{\frac{1}{2}}(a_j + a_l)/(1+\rho)]^2}{1-\rho}\right] dy\right].$$

Let

$$z = \left(\frac{1+\rho}{1-\rho}\right)^{\frac{1}{2}} \frac{y - \rho^{\frac{1}{2}}(a_j + a_l)}{1+\rho},$$

$$\rho'' = \rho/(1+2\rho),$$

$$a_{ijl} = [(1+\rho)\,(1-\rho)^{-1}\,(1+2\rho)^{-1}]^{\frac{1}{2}}\,[a_i - \rho(a_j + a_l)/(1-\rho)] \quad (i = 1, \ldots, n)$$

and $\mathbf{a}_{jl}$ be a $n \times 1$ vector with ith element a_{ijl}. Then $\partial P_0 / \partial \rho$ can be written as

$$= \frac{(2\pi)^{-\frac{1}{2}}}{2} \sum_{l=1}^{n} \sum_{\substack{j=1 \\ j \neq l}}^{n} \phi(a_l, a_j; \rho) \int_{-\infty}^{\infty} \left[\prod_{\substack{i=1 \\ i \neq j, l}}^{n} \Phi\left(\frac{a_{ijl} - (\rho'')^{\frac{1}{2}} z}{(1-\rho'')^{\frac{1}{2}}}\right)\right] \exp\left(-\tfrac{1}{2}z^2\right) dz$$

$$= \sum_{l=1}^{n} \sum_{j>l}^{n} \phi_2(\mathbf{a}, \mathbf{m}; \rho)\, \Phi_2(\mathbf{a}_{jl}, \mathbf{m}; \rho''),$$

where

$$\mathbf{m} = \mathbf{I}(j) + \mathbf{I}(l).$$

Similarly we find

$$\frac{\partial \Phi(\mathbf{a}, \mathbf{k}; \rho)}{\partial \rho} = \sum_{l=1}^{n} \sum_{j>l}^{n} \phi_2(\mathbf{a}, \mathbf{m}; \rho)\, \Phi_2(\mathbf{a}_{jl}, \mathbf{k} + \mathbf{m}; \rho'')\, A_{\mathbf{km}}.$$

If we let

$$\rho' = \rho/(1+\rho),$$

$$b_{ij} = (a_i - \rho a_j)/(1-\rho^2)^{\frac{1}{2}},$$

and $\mathbf{b}_j$ be the $n \times 1$ vector with ith element b_{ij} then

$$\frac{\partial P_{\mathbf{0}}}{\partial a_j} = \phi_2(\mathbf{a}, \mathbf{m}; \rho)\, \Phi_2(\mathbf{b}_j, \mathbf{m}; \rho'),$$

where $\mathbf{m} = \mathbf{I}(j)$.

Also
$$\frac{\partial \Phi_2(\mathbf{a}, \mathbf{k}; \rho)}{\partial a_j} = \phi_2(\mathbf{a}, \mathbf{m}; \rho)\, \Phi_2(\mathbf{b}_j, \mathbf{k}+\mathbf{m}; \rho')\, A_{\mathbf{km}}.$$

Hence the differential of $\Phi_2(\mathbf{a}, \mathbf{k}, \rho)$ with respect to ρ and a_j can be expressed in terms of ordinates of the normal distribution and multinormal integrals. $\Phi_2(\mathbf{a}, \mathbf{k}; \rho)$ is a $n - b(\mathbf{k})$ dimensional integral and so $n - b(\mathbf{k}) - 2$ and $n - b(\mathbf{k}) - 1$ dimensional integrals are used in evaluating the derivatives with respect to ρ and a_j.

When $n = 2$ the derivatives of $P_{\mathbf{0}}$ with respect to ρ and a_1 reduce to

$$\frac{\partial P_{\mathbf{0}}}{\partial \rho} = \phi(a_1, a_2; \rho),$$

$$\frac{\partial P_{\mathbf{0}}}{\partial a_1} = \phi(a_1)\, \Phi\left(\frac{a_2 - \rho a_1}{(1-\rho^2)^{\frac{1}{2}}}\right) \quad \text{as given by Tallis (1962).}$$

28

Estimating the heritability of all-or-none and categorical traits by offspring-parent regression

by

R. Thompson, B.J. McGuirk and A.R. Gilmour

Sonderdruck aus Zeitschrift für Tierzüchtung und Züchtungsbiologie
Bd. 102 (1985), H. 5, S. 342–354
VERLAG PAUL PAREY · HAMBURG 1 · SPITALERSTRASSE 12
Rechte, auch die der Übersetzung, des Nachdrucks, der photomechanischen Wiedergabe und der Speicherung in Datenverarbeitungsanlagen, vorbehalten, © 1985 Verlag Paul Parey, Hamburg und Berlin

Animal Breeding Research Organization, Edinburgh, Scotland;
Division of Animal Production, Blacktown, NSW, Australia;
Department of Agriculture, Agricultural Research and Veterinary Centre, Orange, NSW, Australia

Estimating the heritability of all-or-none and categorical traits by offspring-parent regression

By R. Thompson, B. J. McGuirk and A. R. Gilmour

Ms. received 11. 3. 1985

Introduction

In recent years considerable attention has been given to the estimation of genetic parameters and the prediction of breeding values for characters which are not normally distributed. The stimulus for much of this work is the desire to incorporate information on categorical traits in the evaluation of breeding value. For example, characters such as calving difficulty and type classification are important in both the beef and dairy cattle industries. Survival and disease traits are important in many species and naturally occur or are scored into a few, often two, classes. Consequently, improved statistical procedures for the analysis and interpretation of all-or-none and categorical traits are of interest to many animal breeders.

All-or-none traits pose statistical problems because means and variances are related. This violates one of the assumptions for the analysis of variance. In the animal breeding context, heritability estimates are related to the frequency of the condition (Robertson and Lerner, 1949). This complicates attempts to specify the degree of genetic variation in an all-or-none trait, for a population represented in several herds/flocks/years in which the incidences vary. It also complicates the comparison of heritability estimates from populations in which incidences differ.

These difficulties can sometimes be overcome if we examine liability to a condition, rather than its incidence. Liability is defined as the sum of genetic and environmental factors influencing susceptibility, and is assumed to be normally distributed (see Falconer 1965). Animals with liability above a notional threshold are assumed to exhibit the condition, so that expression is an all-or-none phenomenom. The categorial traits considered here are those representing an ordered level of response, indicating differential susceptibility among the affected animals. They are assumed to arise from multiple thresholds superimposed on a single underlying liability variable.

Under the liability model for all-or-none traits in a single population, the heritability of incidence is $z^2/q\,(1-q)$ times the heritablity of liability, where q is incidence of the character in the population and z is the ordinate of a standard normal distribution at the truncation point above which lies an area q (see for example Robertson and Lerner 1949; Hill and Smith 1977). Simulation studies (Van Vleck 1972; Olausson and Rönningen 1975) have shown that this expression provides a reliable description of the relationship between half-sib heritability estimates for incidence and liability. It is now quite common for researchers to convert heritability estimates to the liability scale using this relationship (see for example Philipsson 1976). Gianola (1979) has extended the expression to multiple thresholds.

U.S. Copyright Clearance Center Code Statement: 0044–3581/85/10205–0342 $ 2.50/0
Z. Tierzüchtg. Züchtgsbiol. 102 (1985) 342–354
© 1985 Verlag Paul Parey, Hamburg und Berlin
ISSN 0044–3581/InterCode: ZTZBAS

However this simple relationship among heritability estimates does not hold in the mixed-model situation where proportions vary between levels of the fixed effects (see GILMOUR, THOMPSON and MCGUIRK, in preparation).

In considering the heritability of a liability, animal breeders have given little attention to offspring-parent estimation procedures. FALCONER (1965) treated such information within the more general framework described above for half-sib estimates. JAMES and MCGUIRK (1982) showed that changes of incidence in both the offspring and parent generations would change heritability estimates for incidence. Varying the offspring incidence has the greater effect. However, JAMES and MCGUIRK (1982) did not indicate how to obtain a direct estimate of the offspring-parent heritability of liability. That is the objective of this paper. After dealing with incidence data, the extension to categorical data will be described. The method of estimation being proposed has the important advantage of being general to mixed model situations. To illustrate the proposed methods, we have made extensive use of data on fleece rot incidence and score in Australian Merino sheep which MCGUIRK and ATKINS (1984) described and for which they presented both half-sib and offspring-parent heritability estimates.

Theoretical development

FALCONER (1965) proposed that characters expressed in an all-or-none fashion might follow the usual laws of quantitative genetic theory if examined at the level of liability, rather than expression. In particular, he proposed that the heritability of liability might be a useful starting point in predicting responses to selection, both in liability and incidence.

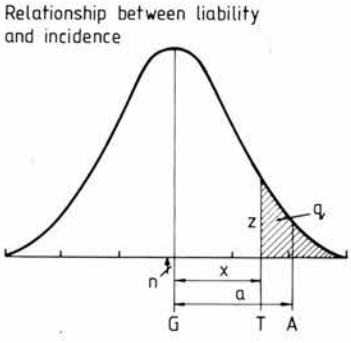


Fig. 1. Relationship between liability and incidence in a population. G = Mean liability of the population; A = Mean liability of affected individuals; q = Proportion of the population affected; x = Deviation of the threshold from the population mean; z = Height of the ordinate at the threshold; a = Mean deviation of affected individuals from the population mean (= z/(1−p))

The following development and nomenclature follows that used by FALCONER (1965), and is pictured in Figure 1. Suppose that the incidence in the parent generation is q. If liability is assumed to be normally distributed with mean G and variance one, the mean liability of affected parents, those that exceed the threshold, T, is A-G = a = z/q, where z is the height of the ordinate at the threshold. The mean liability of unaffected parents is n = −z/(1−q). The difference in mean liability between the two groups is thus z/q(1−q) or z/pq, if p+q = 1.

Let us now assume that the mean liability in the offspring generation is L and consider the liability of offspring of affected and unaffected sires mated to unselected dams. It follows from animal breeding theory that the mean liability of offspring of affected sires (G_1) is

$$L + \frac{1}{2}\, h^2\, \frac{z}{q} \qquad (1)$$

and the mean liability of offspring of unaffected sires (G_0) is

$$L + \frac{1}{2}\, h^2 \left(\frac{-z}{1-q} \right) \qquad (2)$$

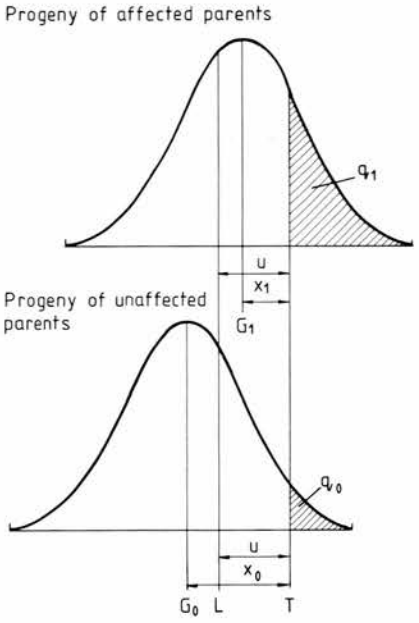


Fig. 2. Two distributions representing liability and incidences in offspring progeny of affected and unaffected parents. L = Mean liability of the population; μ = Deviation of threshold from the population mean; q_1 = incidence among progeny of affected parents; x_1 = Deviation of threshold from the mean liability of progeny of affected parents; G_1 = Mean liability of affected parents; q_0 = Incidence among unaffected parents; x_0 = Deviation of threshold from the mean liability of unaffected parents; G_0 = Mean liability of unaffected parents

This situation is pictured in Figure 2.

If liability is normally distributed the proportion unaffected in these offspring groups, p_i, is related to the difference in liability between the threshold T and the population mean by the expression

$$p_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_i} \exp(-y^2/2)\, dy = \Phi(x_i) \tag{3}$$

Correspondingly, the proportion affected, q_i, is equal to $1 - p_i = \Phi(-x_i)$. The term x_i is commonly referred to as the Normal Equivalent Deviate of p_i.

It follows from (1) and (2) that

$$x_1 = \mu - \frac{1}{2} h^2 \frac{z}{q} \tag{4}$$

$$x_0 = \mu - \frac{1}{2} h^2 \left(\frac{-z}{1-q}\right) \tag{5}$$

The difference in mean liability between offspring of the two parental types is $x_0 - x_1$, which is equal to

$$\frac{1}{2} h^2 \left(\frac{z}{(1-q)} + \frac{z}{q}\right), \text{ so that}$$

$$h^2 = 2(x_0 - x_1)\frac{q(1-q)}{z} \tag{6}$$

Falconer (1965, equation 6) has a similar formula in terms of liability in the control (x_c) and affected (x_r) groups, and the general population incidence.

We can relate the heritability estimate so obtained with the heritability for incidence. If we use the approximate relationship that

$$z^1 (x_0 - x_1) = p_0 - p_1$$

where z^1 is the ordinate corresponding to the mean incidence in the offspring generation, then from (6) above

$$h^2 = \frac{2(p_0 - p_1)}{z^1} \cdot \frac{q(1-q)}{z} \qquad (7)$$

We can think of $q(1-q)/zz^1$ as the appropriate adjustment for transforming heritability to the underlying scale, taking into account different incidences in the parent and offspring generations. As James and McGuirk (1982) pointed out, variation in offspring incidence has the greater effect. When offspring and parent incidences are equal, the scaling factor is the same as for the half-sib estimates, namely that the heritability of liability is $q(1-q)/z^2$ times that for incidence.

Example of calculations

We can illustrate this approach using data described in Appendix 1. The incidence of fleece rot in the parental generation for Cohort 1 was 68.8 per cent, so that the mean liability of affected parents is 0.514 (=z/q) standard deviations and the mean liability of unaffected parents −1.134 (= −z/(1−q). Among the offspring of affected parents, 3 out of 21 (14.3 per cent) were unaffected, compared with 9 out of 41 (22.0 per cent) of the offspring of unaffected parents. Using equations (3) with equations (4) and (5),

$$\frac{3}{21} = 0.143 = \Phi(\mu - \frac{1}{2} h^2 0.514)$$

and

$$\frac{9}{41} = 0.220 = \Phi(\mu - \frac{1}{2} h^2 (-1.134))$$

Using equation (6)

$$\begin{aligned} h^2 &= 2(-0.772 - -1.067)/(1.134 + 0.514) \\ &= 0.590/1.648 \\ &= 0.36 \end{aligned}$$

The heritability of incidence would usually be estimated as $2(p_0 - p_1)$, or 2(0.220–0.143) to give 0.154. Thus the heritability of liability is approximately 2.5 times greater.

This is relatively straightforward for one set of parents and offspring. Where there are more sets, the method needs to be generalised. Let us consider the situation where there are a number of groups of sires (i) which differ in incidence and groups of progeny (j) which differ in incidence. These groups may represent animals born in different years or at different locations. Extending formulae (4) and (5), the Normal Equivalent Deviates are

$$x_{ij1} = \mu_{ij} - \frac{1}{2} h^2 \frac{z_i}{qi}$$

and

$$x_{ijo} = \mu_{ij} - \frac{1}{2} h^2 \frac{z_i}{1 - q_i}$$

where z_i and q_i relate to parental incidences in the i^{th} sire group and μ_{ij} measures the difference between the threshold and the mean liability of the ijth offspring group.

The Normal Equivalent Deviate is thus a linear function of the fixed effects, heritability and mean parental liability. These quantities may then be estimated using generalized linear models (GLM; McCullagh and Nelder 1983) and a model which fits the fixed effects and has mean parental liability as a covariate. Programs such as GENSTAT, GLIM and REG (Gilmour 1983) are specifically designed for such analyses. It is usual to work with $-x_{ijk}$,

the Normal Equivalent Deviate of q_{ijk}, so that the coefficient of the heritability estimate is positive. In the next sections we will estimate the heritability of liability to fleece rot using this approach and compare the estimate with the heritability of fleece rot incidence.

Description of fleece rot data set and previously obtained offspring-parent heritability estimates

Fleece rot is a mild superficial dermatitis which can affect sheep after prolonged rainfall, sufficient to keep the skin wet for 4–5 days. Because of this dependence on rain, incidence varies widely between years. In this data set, fleece rot was recorded in hogget (16 months of age) rams and ewes in two unselected control flocks at Trangie Agricultural Research Centre, N.S.W. (see Mc Guirk and Atkins 1984 for further details). Each sheep was given a score of from 0, indicating no evidence of fleece rot, to 5, very severe fleece rot lesions. The data were collected over a 15 year period. Different scorers were involved, and the scoring was entirely subjective, with scores being allocated without the aid of photographic or any

Table 1. **Summary of data used in offspring-parent heritability analyses**

Year of Birth	Season	No. of Animals scored	Incidence (%)	Average Score	Included in Offspring-Sire analyses as		Included in Offspring-Dam analyses as	
					Off-spring	Sire	Off-spring	Dam
1961	1	48	89.6	2.52				*
1962	1	80	52.5	1.14		*		*
1962	2	80	68.8	1.30		*		*
1963	1	92	84.8	2.08				*
1963	2	72	68.1	1.32				*
1964	1	82	87.9	1.85	*			*
1964	2	68	77.9	1.49	*		*	*
1967	1	69	27.5	0.33		*	*	*
1967	2	72	18.1	0.26		*	*	*
1968	1	70	27.1	0.53		*	*	*
1968	2	80	38.8	0.83		*	*	*
1969	1	85	22.4	0.29	*		*	*
1969	2	62	22.5	0.35	*		*	*
1970	1	62	25.2	0.40	*	*	*	*
1970	2	143	21.0	0.37	*		*	*
1971	1	77	13.2	0.22	*	*	*	*
1971	2	123	16.3	0.30	*		*	*
1972	1	73	57.5	0.87	*	*	*	*
1972	2	75	45.3	0.83		*	*	*
1973	1	54	7.4	0.17	*		*	*
1973	2	49	8.2	0.14			*	
1974	1	264	38.3	0.84	*		*	
1975	1	327	10.7	0.21			*	
					(11)	(10)	(17)	(20)

other standards. In all, 2207 animals were scored, and the percentages scored as 0, 1, 2, 3, 4 and 5 respectively were 62.0, 15.3, 11.1, 7.4, 3.4 and 0.8 per cent.

The average incidence was approximately 40 %, but it ranged from 7 to 90 % among years. The sheep were born in 23 year*season groups; the incidence and average score in each of these is shown in Table 1. The term "cohort" will be used to define a group of sheep born in a particular year and season. McGuirk and Atkins (1984) concluded that cohort was the only environmental factor with a statistically significant effect on fleece rot incidence or score. Table 1 also indicates those cohorts involved in the offspring-dam and offspring-sire analyses, and whether they contributed offspring or parents.

In the offspring-parent analyses, parents usually had offspring born in more than one year. To account for cohort effects in both offspring and parents, McGuirk and Atkins (1984) generated a joint cohort-of-offspring * cohort-of-parent main effect, and this approach is also used here. However, the original data set has been reduced to include only those joint cohorts which contain offspring from both affected and unaffected parents. There were 13 such cohorts in the offspring-sire analyses, with 892 pairs of records. There were 1118 pairs of records in the offspring-dam set, representing 68 joint cohorts.

A summary of the least squares analyses conducted on the offspring-sire and offspring-dam data sets is given in Table 2. After adjusting for cohort effects, parent fleece rot score was fitted as a linear covariate, and the regressions of offspring incidence and score on parent score estimated. The heritability of score was calculated as twice the regression of offspring score on parent score. The heritability of incidence was calculated as twice the difference in offspring incidence between parents classed as unaffected and affected.

Table 2. **Summary of offspring-parent analyses of fleece rot score and percentage incidence using least squares methods** (Standard errors in parenthesis)

	Offspring-sire Analyses	Offspring-dam Analyses
1. Regression of offspring score on parent score	.113 (±.031)	.107 (±.027)
2. Regression of offspring (%) incidence on parent score	3.9 (±1.3)	3.5 (±1.2)
3. Fleece rot (%) incidence in progeny of		
(a) Unaffected parents	30.5 (±1.9)	24.9 (±2.1)
(b) Affected parents	39.1 (±2.2)	32.0 (±2.3)

The heritability of fleece rot score was 0.23 (±.06) in the offspring-sire data and 0.21 (±.05) in the offspring-dam data. Mean incidence increased as parent score increased, indicating that the parental scores reflected differential genetic predisposition to fleece rot. The heritability estimates for incidence were 0.17(±.05) and 0.14(±.05) in the offspring-sire and offspring-dam data sets respectively.

Direct estimates of offspring-parent regression of liability to fleece-rot

We now illustrate the method by estimating the heritability of liability to fleece rot. The data set is the same one as described in the previous section.

Four items need to be specified to the programs that fit general-ized linear models. For our example they are

1. The dependent variable. This is AFF, the number affected out of N, the total number scored.
2. The independent variables. These include the factor COHORT and a variate LM which contains the mean liability of the parental groups (unaffected and affected). The regression coefficient associated with COHORT and LM measure the cohort effects (μ_{ij}) and $h^2/_2$ respectively.
3. The error distribution. The numbers affected (AFF) out of N are assumed to follow the binomial distribution.
4. A link function. This converts the linear predictor x_{ij}, a function of μ_{ij}, $h^2/_2$, COHORT and LM, to the proportion affected (AFF/N). The probit transformation, $p_i = \Phi(x_i)$, is used to change from the underlying (x) scale to the incidence (p) scale.
   Data can be prepared as shown in Appendix 1. For each cohort, two records are required, one for the offspring of unaffected parents, the other for the offspring of affected parents. In each we specify N, AFF and LM. The information on offspring and parental incidences and the Normal Equivalent Deviate (NED) of the offspring incidence, are not required as data inputs, but are included in Appendix 1 to help explain the method of analysis.

1. Offspring-sire analyses

The offspring-sire heritability estimate for liability, calculated as twice the within-cohort regression of offspring liability on parent liability, was 0.36(±0.11). This can be compared with the offspring-sire estimate for incidence, calculated from the regression of offspring incidence on parent incidence on the same data, of 0.17(±.05).

The differences in the two estimation procedures are illustrated in Figures 3a and 3b for the sample of five offspring-sire cohort groups described in Appendix 1. These represent a range of offspring incidences. Heritability is estimated as twice the average slope of the lines.

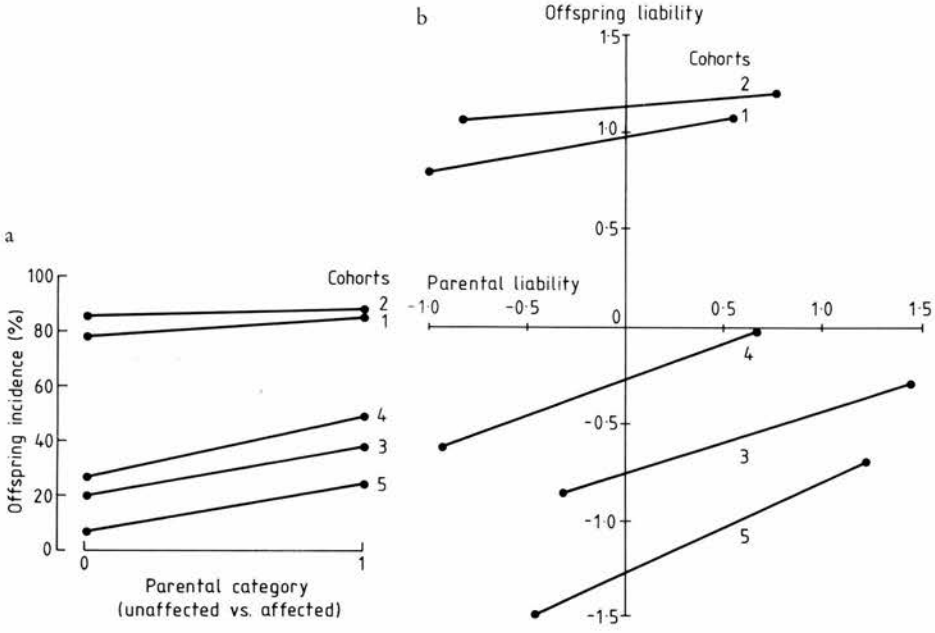


Fig. 3. The relationship between fleece rot susceptibility in offspring and parents in five cohort groups. (a) Incidence; (b) Liability

In figure 3a, offspring incidences are plotted against their parental scores coded as 0 or 1. In Figure 3b, the Normal Equivalent Deviates of the offspring incidences are plotted against their parent's mean liability. Both axes are thereby stretched. A difference in offspring incidence of 7.7 per cent, 0.077, in Cohort 1 (78.0 v 85.7 per cent) is now equivalent to a difference in liability of 0.295 (1.067–0.772). The stretching factor is equal to $1/z^1$, which for a mean offspring incidence of 81.8 %, is equal to 1/.264, or 3.78. Similarly, for a parental incidence of 68.8 per cent, the stretching factor (z/pq) is equal to 1.648. As this is less than the stretching on the offspring scale, the heritability of liability is greater than that for incidence. For Cohort 1, the heritability on the liability scale is equal to $2 \times 0.295/1.648$, or 0.36, compared with an estimate for incidence of 2×0.077, or 0.154.

To examine the separate effects of changing the parental and offspring scales from incidence to liability, we carried out the analyses described in Table 3.

Table 3. **Regression coefficients (and standard errors) for offspring-sire analyses, comparing incidence and liability scales for both offspring and sire, after adjusting for a combined offspring/sire cohort effect.** Residual Deviances$^+$ estimated with 12df

	Sire Scale	
	Incidence	Liability
Offspring Incidence	0.075(±.031)	0.046(±.019)
	15.72$^+$	15.50$^+$
Offspring Liability	0.291(±.096)	0.180(±.059)
	13.99$^+$	13.65$^+$

The regression coefficients are on different scales. Changing from the sire incidence to liability scale multiplies the regression coefficients by 0.61 (Offspring Incidence analysis) and 0.62 (Offspring Liability). These values may be compared with 0.66 obtained when values of pq/z for each of the 13 cohorts are averaged. Changing from the offspring incidence to liability scale multiplies the regression coefficients by 3.88 (Parent Incidence) and 3.91 (Parent Liability). These values may be compared with 3.91 which is the average value of $1/z^1$ when weighted by the numbers of progeny in each cohort. It is intermediate between the simple average of the values for $1/z^1$ (= 4.10) and the value of $1/z^1$ calculated for the average incidence of 36 per cent (= 2.7).

As anticipated, the heritability estimate for liability was higher than that for incidence (0.36 v. 0.15), in the ratio here of 2.4:1. We had expected the Residual Deviance would be lower for analyses on the liability scale, as it measures the homogeneity of regression coefficients among the cohorts. However there was little to choose between the models on this ground.

2. *Offspring-dam analyses*

The offspring-dam heritability of liability to fleece rot (Table 4) was 0.35(±.12), compared with an estimate of 0.22(±.05) for incidence, a ratio of 1.59:1. The Residual Deviances were not lower for analyses on the liability scale.

The effects of changing the parent and offspring scales from incidence to liability are as expected. Changing from the dam incidence to the liability scale multiplies the regression coefficients by 0.54 (Offspring incidence) and 0.58 (Offspring liability). Changing from the offspring incidence to liability scale increases the regression coefficients by a factor of 2.80 (parent incidence analysis) and 2.97 (parent liability).

Table 4. **Regression coefficients (and standard errors) for offspring-dam analyses, comparing incidence and liability scales for offspring and dams, after adjusting for a combined dam/offspring cohort effect.** Residual Deviances+ estimated with 67 d.f

	Dam Scale	
	Incidence	Liability
Offspring Incidence	0.108(±.027)	0.059(±.016)
	86.73+	87.36+
Offspring Liability	0.303(±.103)	0.175(±.060)
	93.94+	94.17+

In the analyses reported in Table 4 we estimated only one fixed effect, which jointly accounted for the cohort effects of both dam and offspring. This was the approach previously adopted by McGuirk and Atkins (1984). The more usual approach would be to fit year of birth of dam and year of birth of offspring as separate main effects, accounting respectively for 19 and 16 degrees of freedom. The remaining 32 degrees of freedom (67 − 19 − 16) might be attributed to interaction effects between the two. This reduced model was fitted for both incidence and liability using REG (Gilmour, 1983). The interaction effects were tested by difference from the full model and were not significant ($P > 0.05$). In none of the four analyses was the dam cohort effect significant and it was deleted from the model, leaving only offspring cohort and dam liability as the independent variables. The regression coefficients and Residual Deviances for these models are given in Table 5.

While the regression estimates in Tables 4 and 5 are similar, the standard errors are lower in Table 5. By removing the non-significant effects of year of birth of the dam and the interaction effect from the model to the error we have improved the precision with which the regression coefficients were estimated.

Table 5. **Regression coefficients (and standard errors) for offspring-dam analyses, comparing incidence and liability scales for both dams and offspring, after adjusting only for offspring cohorts.** Residual Deviances+ estimated with 118 df.

	Dam Scale	
	Incidence	Liability
Offspring Incidence	.103 (±.026)	.048 (±.016)
	142.43+	148.02+
Offspring Liability	.346 (±.093)	.174 (±.057)
	142.81+	147.41+

The heritability estimates for liability and incidence are similar to those described in Table 4, namely 0.35 (±.11) and 0.20 (±.05). The observed stretching factors in changing from offspring incidence to liability are now 3.4 (dam incidence) and 3.6 (dam liability) and are close to the value of 3.9 obtained from equation (7). Looking at the Residual Deviances, the regression estimates of offspring liability on dam liability are less homogeneous than for the other three scale combinations. Some of the Residual Deviances exceed the 5 % chi-square value for 118 df of 144.3 which suggests that important sources of variation are missing from the model. For characters which are heritable, sire effects would be an obvious candidate.

Extension to multiple threshold analysis

The procedure described for an all-or-none character extends directly to multiple threshold categorical traits. The mean liability of parents in the k^{th} score category in a particular parental group may be estimated as

$$\frac{z_{k-1}-z_k}{p_k}$$

where z_{k-1} and z_k are the ordinates at the $(k-1)^{th}$ and k^{th} thresholds, and p_k is the proportion in the k^{th} category. The multiple threshold analysis is more complex because the proportions in each category are correlated. This difficulty is accommodated by the use of composite link functions (THOMPSON and BAKER 1981) and the appropriate analysis can be conducted using GENSTAT, GLIM or REG. The analysis does assume that the variance of liability is constant in the different levels of the fixed effects, and that the thresholds also remain fixed. Given these assumptions, the proportions in each category are solely a function of the mean liability.

This method of analysis was used to estimate the heritability of liability to fleece rot, from the fleece rot scores in the offspring-sire data set. In these analyses, the very few animals with scores 4 and 5 were recoded as 3s. As in the all-or-none analyses, a cohort effect with 12 degrees of freedom was used to account jointly for cohort effects in both sire and offspring generations. The heritability of fleece rot score was recalculated and an estimate of 0.25(±.06) obtained; the value was 0.23 (see Table 2), when the higher order scores were not pooled.

A subset of the data used to estimate the heritability of liability is illustrated in Appendix 2. In Cohort Group 7 there were animals in the parental cohort with a score of 2, but no sires with score 2 were used. This is the reason for the row of zero values in offspring categories for this sire score.

The regression estimates for offspring liability on both sire liability and score are summarised in Table 6. Fitting sire liability accounted for more of the variation than did sire score, hence the lower Residual Deviance. Again sire effects would have contributed to the significant Residual Deviances. The heritability of liability, twice the regression of offspring liability on sire liability, was 0.40(±.10), which is similar to the heritability estimate for liability obtained from the all-or-none classification (see Table 3).

Table 6. **Regression coefficients (and standard errors) for offspring liability on either sire score or liability, adjusted for a combined sire/offspring cohort effect.**
Residual Deviances+ estimated with 101 df

	Sire Scale	
	Score	Liability
Offspring Liability	0.158 (±.042)	.201 (±.049)
	188.67*†	185.68*†

* P < 0.05

Discussion

In this paper we have described a method for estimating the heritability of liability directly from offspring-parent information on incidence, where the character is coded as two or more ordered categories. The method can be seen as an extension of FALCONER's (1965)

proposal for estimating the heritability of liability to mixed model analysis. Such an extension is necessary when mean incidences and scores vary between levels of a fixed effect. The method does suffer from the same deficiences as Falconer's method. For example, no allowance is made for the reduction in variance among progeny of parents from particular categories, or the fact that liability among such a group of progeny will no longer be normally distributed (see for example REICH, JAMES and MORRIS 1972). The former effect is likely to be more important and will mean that heritability is slightly underestimated (see REICH, JAMES and MORRIS 1972).

The approach described has a number of other applications. It could be used to estimate the repeatability of liability, from the regression of subsequent on early performance (LUSH 1956; RUTLEDGE 1977). It could also be used in offspring-parent analyses to estimate genetic regressions among pairs of scored characters or among scored and metric traits. In the latter situation if the scored trait was measured on the parent generation, then parental liability would be included as a covariate. If the scored trait was on the offspring, then a probit link function would be used.

For the analysis of all-or-none traits, there must be observations in each of the four parent/offspring categories; there must be some affected and unaffected parents with both affected and unaffected offspring. When fixed effects are included, for example parental cohorts, it is not necessary that each of the four categories be represented in each level of the fixed effect, that is in every parent cohort. However the parental cohort will not contribute information to offspring-parent regression unless three of the categories are represented. This of necessity means that among the animals from a particular cohort actually used as parents, there must be affected and unaffected animals. For categorical traits, both parents and offspring should be represented in at least two classes within each cohort.

The method of analysis described requires that we know the mean liability of parents in the different categories. However, these mean liabilities are not known and the precision with which they are estimated will depend on the total numbers scored and the distribution of scores. The heritability estimates may also be biassed downwards when samples are small and incidences are extreme. As an example of the possible bias, if the true incidence is six per cent and the total number scored into affected and unaffected classes is 20, then the bias is of the order of two per cent. This bias is unlikely to be a problem in situations likely to be encountered in practice.

The analyses of fleece rot data indicate considerable genetic variation in liability. The three heritability estimates obtained for liability were very similar; $0.36(\pm.11)$ and $0.40(\pm.10)$ respectively for the incidence and score data in the offspring-sire data set and $0.35(\pm.11)$ for offspring-dam incidences. These estimates are similar to the half-sib estimate for liability in this data set (GILMOUR, THOMPSON and MCGUIRK, in preparation).

Acknowledgements

We are grateful to the Director of Research, Trangie Agricultural Research Centre, for permission to use the data on fleece rot referred to in the paper. This study was undertaken during a visit to Australia by ROBIN THOMPSON, a visit made possible by travel grants from CSIRO's Divisions of Animal Production and Mathematics and Statistics. We thank the staff of the New South Wales Department of Agriculture's Biometrical Branch for their assistance in computing, and Mr. M. BLAIR, CSIRO Prospect, for preparing the figures.

Appendix 1

A sample of all-or-none data examined to estimate the offspring-parent heritability to fleece rot

		OFFSPRING				PARENTS	
	COHORT	N	AFF	(Incidence) %	(NED)	LM	(Incidence) %
(Unaffected parents)	1	41	32	78.0	0.772	-1.134	68.8
(Affected parents)	1	21	18	85.7	1.067	0.514	
(Unaffected parents)	2	21	18	85.7	1.067	-0.838	52.5
(Affected parents)	2	61	54	88.5	1.200	0.758	
(Unaffected parents)	3	46	9	19.6	-0.856	-0.321	18.1
(Affected parents)	3	13	5	38.5	-0.292	1.455	
(Unaffected parents)	4	71	19	26.8	-0.619	-0.922	57.5
(Affected parents)	4	63	31	49.2	-0.020	0.681	
(Unaffected parents)	5	14	1	6.8	-1.491	-0.454	27.1
(Affected parents)	5	37	9	24.3	-0.697	1.222	

Appendix 2

An example of a subset of data on fleece rot scores used to estimate the heritability of liability

COHORT GROUP	PARENTAL SCORE	(PROPORTION IN COHORT)	PARENTAL LIABILITY	DISTRIBUTION OF OFFSPRING SCORES				
				0	1	2	3	TOTAL
6	0	0.475	-0.838	3	7	3	8	21
6	1	0.175	0.159	1	6	3	7	17
6	2	0.163	0.623	6	12	7	8	33
6	3	0.187	1.436	0	7	1	3	11
7	0	0.612	-0.625	24	2	0	1	27
7	1	0.188	0.549	15	0	0	1	16
7	2	0.037	0.911	0	0	0	0	0
7	3	0.163	1.513	11	0	0	1	12

Summary

The estimation of the heritability of liability for a binomial or multiple threshold categorical trait is described and demonstrated. The method is an extension of that proposed by FALCONER (1965) and allows for fixed effects in both the parent and offspring generations. The probit transformation is used to relate the liability scale to observations on the probability scale. Data on fleece rot incidence and scores in Merino sheep have been used to illustrate the method.

Résumé

Estimation de l'héritabilité de caractéristiques catégoriques par la regression descendance-parents

L'estimation de l'héritabilité de la responsabilité relative à un caractère binomial catégorique ou d'un caractère à seuils multiples est décrite et démontrée. La méthode est une extension de celle proposée par FALCONER (1965) et permet la considération d'effets fixes dans les générations des parents aussi bien que dans les générations des descendants. La transformation «probit» est utilisée pour mettre en relation l'échelle de responsabilité aux observations sur l'échelle de probabilité.

Les dates d'incident de putréfaction de toison et points en moutons mérino ont été utilisées pour l'illustration de la méthode.

Resumen

Estimación de heredabilidad de características binomiales y categoricas por la regresión de la descendencia sobre los padres

Se describe y demuestra, la estimación de la heredabilidad de propensión, para una característica binomial o de umbral múltiple. El método es una extensión del propuesto por FALCONER (1965) y toma en cuenta los efectos fijos en ambos, los padres y de la descendencia. La transformación «probit», es usada para relacionar la escala de exposición por observación, sobre la escala de probabilidad. Datos sobre incidencia de vellón podrido (en puntos) en ovejas Merino, han sido usadas para ilustrar el método.

Zusammenfassung

Schätzung der Heritabilität kategorischer Merkmale durch Nachkommen-Elternregression

Die Schätzung der Heritabilität der Anfälligkeit bezüglich eines binomialen kategorischen Merkmales und eines solchen mit mehrfachen Schwellen wird beschrieben und demonstriert. Die Methode ist eine Fortentwicklung der von FALCONER (1965) vorgeschlagenen und gestattet die Berücksichtigung fixer Wirkungen sowohl in den Eltern wie auch in den Nachkommengenerationen. Die Probittransformation wird verwendet, um die Verbindung zwischen der zugrundeliegenden Skala der Anfälligkeit und der Beobachtungen auf der Wahrscheinlichkeitsskala herzustellen.

Es werden Unterlagen von Vliessfäulebefall und -punkten in Merinoschafen zur Illustration der Methode verwendet.

References

FALCONER, D. S., 1965: The inheritance of liability to certain diseases, estimated from the incidence among relatives. Annals. of Human Genetics, **29:** 51–76.

GIANOLA, D., 1979: Heritability of a polychotomous characters. Genetics **93:** 1051–1055.

GILMOUR, A. R., 1983: REG – A generalized linear models programme. Miscellaneous Bulletin 1, New South Wales Department of Agriculture, 120 pp.

HILL, W. G.; SMITH, C., 1977: Estimating 'heritability' of a dichotomous trait. Biometrics **33:** 231–236.

JAMES, J. W.; MCGUIRK, B. J., 1982: Regression of offspring on parent for all-or-none trait. Z. Tierzüchtg. und Züchtgsbiol. **99:** 308–314.

LUSH, J. L., 1956: "Answer to Query". Biometrics **12:** 84–88.

MCCULLAGH, P.; NELDER, J. A., 1983: Generalized Linear Models. Chapman and Hall; London.

MCGUIRK, B. J.; ATKINS, K. D., 1984: Fleece rot in Merino sheep: 1. The inheritance of fleece rot in a flock of medium-wool Peppin Merinos. Austr. J. Agric. Res. **35:** 425–434.

OLAUSSON, A.; RONNINGEN, K., 1975. Estimation of genetic parameters for threshold characters. Acta. Agriculturae Scandanavica **25:** 201–208.

PHILIPSSON, J., 1976: Studies on calving difficulties, stillbirth and associated factors in Swedish cattle breeds. Acta Agriculturae Scandinavica **26:** 151–164.

ROBERTSON, A.; LERNER, I. M., 1949: The heritability of all-or-none traits: variability of poultry. Genetics **34:** 395–411.

REICH, T.; JAMES, J. W.; MORRIS, C. A., 1972: The use of multiple thresholds in determining the mode of transmission of semicontinuous traits. Animals of Human Genetics **36:** 163–184.

RUTLEDGE, J. J., 1977: Repeatability of threshold traits. Biometrics **33:** 395–399.

THOMPSON, R.; BAKER, R. J., 1981: Composite link functions in generalized linear models. Applied Statistics **30:** 125–131.

VAN VLECK, L. D., 1972: Estimation of heritability of threshold characters. Journal of Dairy Science **55:** 218–225.

Authors' addresses: *Mr. R.* THOMPSON, Animal Breeding Research Organisation, King's Buildings, West Mains Road, Edinburgh EH9 3JH, Scotland; Dr. B. J. MCGUIRK, CSIRO, Division of Animal Production, PO Box 239, Blacktown, NSW 2148, Australia; Dr. A. R. GILMOUR, NSW Department of Agriculture, Agricultural Research and Veterinary Centre, Forest Road, Orange, NSW 2800, Australia

29

Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty, birth weight and pelvic opening

by

J.L. Foulley, D. Gianola and R. Thompson

Génét. Sél. Evol., 1983, **15** (3), 401-424

Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty, birth weight and pelvic opening

J.L. FOULLEY, D. GIANOLA* and R. THOMPSON**

I.N.R.A., Station de Génétique quantitative et appliquée, Centre de Recherches zootechniques, F 78350 Jouy-en-Josas
**Department of Animal Science, University of Illinois, Urbana, Illinois 61801, U.S.A.,*
***A.R.C. Unit of Statistics, University of Edinburgh, Mayfield Road, Edinburgh, EH9 3JZ, Scotland*

Summary

A method of prediction of genetic merit from jointly distributed quantal and quantitative responses is described. The probability of response in one of two mutually exclusive and exhaustive categories is modeled as a non-linear function of classification and « risk » variables. Inferences are made from the mode of a posterior distribution resulting from the combination of a multivariate normal density, a priori, and a product binomial likelihood function. Parameter estimates are obtained with the Newton-Raphson algorithm, which yields a system similar to the mixed model equations. « Nested » Gauss-Seidel and conjugate gradient procedures are suggested to proceed from one iterate to the next in large problems. A possible method for estimating multivariate variance (covariance) components involving, jointly, the categorical and quantitative variates is presented. The method was applied to prediction of calving difficulty as a binary variable with birth weight and pelvic opening as « risk » variables in a *Blonde d'Aquitaine* population.

Key-words : sire evaluation, categorical data, non-linear models, prediction, Bayesian methods.

Résumé

Prédiction génétique à partir de données binaires et continues : application aux difficultés de vêlage, poids à la naissance et ouverture pelvienne.

Cet article présente une méthode de prédiction de la valeur génétique à partir d'observations quantitatives et qualitatives. La probabilité de réponse selon l'une des deux modalités exclusives et exhaustives envisagées est exprimée comme une fonction non linéaire d'effets de facteurs d'incidence et de variables de risque. L'inférence statistique repose sur le mode de la distribution a posteriori qui combine une densité multinormale a priori et une fonction de vraisemblance produit de binomiales. Les estimations sont calculées à partir de l'algorithme de Newton-Raphson qui conduit à un système d'équations similaires à celles du modèle mixte. Pour les gros fichiers, on suggère des méthodes itératives de résolution telles que celles de Gauss-Seidel et du gradient conjugué. On propose également une méthode d'estimation des composantes de variances et covariances relatives aux variables discrètes et continues. Enfin, la méthodologie présentée est illustrée par une application numérique qui a trait à la prédiction des difficultés de vêlage en race bovine *Blonde d'Aquitaine* utilisant d'une part, l'appréciation tout-ou-rien du caractère, et d'autre part, le poids à la naissance du veau et l'ouverture pelvienne de la mère comme des variables de risque.

Mots-clés : Évaluation des reproducteurs, données discrètes, modèle non linéaire, prédiction, méthode bayesienne.

I. Introduction

In many animal breeding applications, the data comprise observations on one or more quantitative variates and on categorical responses. The probability of « successful » outcome of the discrete variate, e.g., survival, may be a non-linear function of genetic and non-genetic variables (sire, breed, herd-year) and may also depend on quantitative response variates. A possible course of action in the analysis of this type of data might be to carry out a multiple-trait evaluation regarding the discrete trait as if it were continuous, and then utilizing available linear methodology (HENDERSON, 1973). Further, the model for the discrete trait should allow for the effects of the quantitative variates. In addition to the problems of describing discrete variation with linear models (COX, 1970; THOMPSON, 1979; GIANOLA, 1980), the presence of stochastic « regressors » in the model introduces a complexity which animal breeding theory has not addressed.

This paper describes a method of analysis for this type of data based on a Bayesian approach; hence, the distinction between « fixed » and « random » variables is circumvented. General aspects of the method of inference are described in detail to facilitate comprehension of subsequent developments. An estimation algorithm is developed, and we consider some approximations for posterior inference and fit of the model. A method is proposed to estimate jointly the components of variance and covariance involving the quantitative and the categorical variates. Finally, procedures are illustrated with a data set pertaining to calving difficulty (categorical), birth weight and pelvic opening.

II. Method of inference : general aspects

Suppose the available data pertain to three random variables: two quantitative (e.g., calf's birth weight and dam's pelvic opening) and one binary (e.g., easy vs. difficult calving). Let the data for birth weight and dam's pelvic opening be represented by the vectors $\mathbf{y}_1$ and $\mathbf{y}_2$, respectively. Those for calving difficulty are represented by a set $\mathbf{Y}$ of indicator variables describing the configuration of the following $s \times 2$ contingency table:

Row	*Category of response*	
	Easy calving	*Difficult calving*
1	n_{11}	$n_{1.} - n_{11}$
2	n_{21}	$n_{2.} - n_{21}$
:	:	:
i	n_{i1}	$n_{i.} - n_{i1}$
:	:	:
s	n_{s1}	$n_{s.} - n_{s1}$

where the s rows indicate conditions affecting individual or grouped records. The two categories of response are mutually exclusive and exhaustive, and the number of observations in each row, $n_{i.} \neq 0$, is assumed fixed. The random quantity n_{i1} (or, conversely, $n_{i.} - n_{i1}$) can be null, so contingency tables where $n_{i.} = 1$, for i = 1, ..., s, are allowed. The data can be represented symbolically by the vector $\mathbf{Y}' = (Y_1, Y_2, ..., Y_s)$, where $Y_i = \sum_{r=1}^{n_i} Y_{ir}$ with Y_{ir} being an indicator variable equal to 1 if a response occurs and zero otherwise.

The data $\mathbf{Y}$, $\mathbf{y}_1$ and $\mathbf{y}_2$, and a parameter vector $\boldsymbol{\theta}$ are assumed to have a joint density $f(\mathbf{Y}, \mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\theta})$ written as

$$f(\mathbf{Y}, \mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\theta}) = f_2(\mathbf{Y}, \mathbf{y}_1, \mathbf{y}_2 | \boldsymbol{\theta}) \cdot f_1(\boldsymbol{\theta}). \tag{1}$$

where $f_1(\boldsymbol{\theta})$ is the marginal or a priori density of $\boldsymbol{\theta}$. From (1)

$$f_4(\boldsymbol{\theta} | \mathbf{Y}, \mathbf{y}_1, \mathbf{y}_2) = f_2(\mathbf{Y}, \mathbf{y}_1, \mathbf{y}_2 | \boldsymbol{\theta}) . f_1(\boldsymbol{\theta}) / f_3(\mathbf{Y}, \mathbf{y}_1, \mathbf{y}_2) \tag{2}$$

where $f_3(\mathbf{Y}, \mathbf{y}_1, \mathbf{y}_2)$ is the marginal density of the data, i.e., with $\boldsymbol{\theta}$ integrated out, and $f_4(\boldsymbol{\theta} | \mathbf{Y}_1, \mathbf{y}_1, \mathbf{y}_2)$ is the a posteriori density of $\boldsymbol{\theta}$. As $f_3(\mathbf{Y}, \mathbf{y}_1, \mathbf{y}_2)$ does not depend on $\boldsymbol{\theta}$, one can write (2) as

$$f_4(\boldsymbol{\theta} | \mathbf{Y}, \mathbf{y}_1, \mathbf{y}_2) \propto f_2(\mathbf{Y}, \mathbf{y}_1, \mathbf{y}_2 | \boldsymbol{\theta}) . f_1(\boldsymbol{\theta}) \tag{3}$$

which is Bayes theorem in the context of our setting. Equation (3) states that inferences can be made a posteriori by combining prior information with data translated to the posterior density via the likelihood function $f_2(\mathbf{Y}, \mathbf{y}_1, \mathbf{y}_2 | \boldsymbol{\theta})$. The dispersion of $\boldsymbol{\theta}$ reflects the a priori relative uncertainty about $\boldsymbol{\theta}$, this based on the results of previous data or experiments. If a new experiment is conducted, new data are combined with the prior density to yield the posterior. In turn, this becomes the a priori density for further experiments. In this form, continued iteration with (3) illustrates the process of knowledge accumulation (CORNFIELD, 1969). Comprehensive discussions of the merits, philosophy and limitations of Bayesian inference have been presented by CORNFIELD (1969), and LINDLEY & SMITH (1972). The latter argued in the context of linear models that (3) leads to estimates which may be substantially improved from those arising in the method of least-squares. Equation (3) is taken in this paper as a point of departure for a method of estimation similar to the one used in early developments of mixed model prediction (HENDERSON *et al.*, 1959). Best linear unbiased predictors could also be derived following Bayesian considerations (RÖNNINGEN, 1971; DEMPFLE, 1977).

The Bayes estimator of $\boldsymbol{\theta}$ is the vector $\hat{\boldsymbol{\theta}}$ minimizing the expected a posteriori risk

$$R(\hat{\boldsymbol{\theta}}; \mathbf{Y}, \mathbf{y}_1, \mathbf{y}_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} l(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) f_4(\boldsymbol{\theta} | \mathbf{Y}_1, \mathbf{y}_1, \mathbf{y}_2) d(\boldsymbol{\theta}) \tag{4}$$

where $l(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$ is a loss function (MOOD & GRAYBILL, 1963). If the loss is quadratic

$$l(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \sum_{i=1}^{k} (\hat{\theta}_i - \theta_i)^2 = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \tag{5}$$

then

$$\frac{\partial R(\hat{\boldsymbol{\theta}}; \mathbf{Y}, \mathbf{y}_1, \mathbf{y}_2)}{\partial \hat{\boldsymbol{\theta}}} = \frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \left\{ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\hat{\boldsymbol{\theta}}' \hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}' \boldsymbol{\theta} - \boldsymbol{\theta}' \hat{\boldsymbol{\theta}} + \boldsymbol{\theta}' \boldsymbol{\theta}] f_4(\boldsymbol{\theta} | \mathbf{Y}, \mathbf{y}_1, \mathbf{y}_2) d\boldsymbol{\theta} \right\}$$

$$= 2[\hat{\boldsymbol{\theta}} - E(\boldsymbol{\theta} | \mathbf{Y}, \mathbf{y}_1, \mathbf{y}_2)] \tag{6}$$

Equating (6) to zero, yields $\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta} | \mathbf{Y}, \mathbf{y}_1, \mathbf{y}_2)$. Note that differentiating (6) with respect to $\hat{\boldsymbol{\theta}}$ yields a positive number, i.e., $\hat{\boldsymbol{\theta}}$ minimizes the expected posterior risk, and $\hat{\boldsymbol{\theta}}$ is identical to the best predictor of $\boldsymbol{\theta}$ in the squared-error sense of HENDERSON (1973). Unfortunately, calculating $\hat{\boldsymbol{\theta}}$ requires deriving the conditional density of $\boldsymbol{\theta}$ given $\mathbf{Y}$, $\mathbf{y}_1$ and $\mathbf{y}_2$, and then computing the conditional expectation. In practice, this is difficult or impossible to execute as discussed by HENDERSON (1973). In view of these difficulties, LINDLEY & SMITH (1972) have suggested to approximate the posterior mean by the mode of the posterior density; if the posterior is unimodal and approximately symmetric,

its mode will be close to the mean. HARVILLE (1977) has pointed out, that if an improper prior is used in place of the « true » prior, the posterior mode has the advantage over the posterior mean, of being less sensitive to the tails of the posterior density.

In (3), it is convenient to write

$$f_2(\mathbf{Y}, y_1, y_2|\boldsymbol{\theta}) = f_6(\mathbf{Y}|y_1, y_2, \boldsymbol{\theta}).f_5(y_1, y_2|\boldsymbol{\theta}) \tag{7}$$

so the log of the posterior density can be written as

$$\ln[f_4(\boldsymbol{\theta}|\mathbf{Y}, y_1, y_2)] = \ln[f_6(\mathbf{Y}|y_1, y_2, \boldsymbol{\theta})] + \ln[f_5(y_1, y_2|\boldsymbol{\theta})] + \ln[f_1(\boldsymbol{\theta})] + \text{const.} \tag{8}$$

III. Model

A. *Categorical variate*

The probability of response (e.g., easy calving) for the i^{th} row of the contingency table can be written as some cumulative distribution function with an argument peculiar to this row. Possibilities (GIANOLA & FOULLEY, 1983) are the standard normal and logistic distribution functions. In the first case, the probability of response is

$$P_{i1} = \Phi(\mu_i) = \int_{-\infty}^{\mu_i} \phi(x)dx \tag{9}$$

where $\phi(.)$ and $\Phi(.)$ are the density and distribution functions of a standard normal variate, respectively, and μ_i is a location variable. In the logistic case,

$$P_{i1} = [1 + e^{-\mu_i^*}] \tag{10}$$

The justification of (9) and (10) is that they provide a liaison with the classical threshold model (DEMPSTER & LERNER, 1950; GIANOLA, 1982). If an easy calving occurs whenever the realized value of an underlying normal variable, $z \sim N(\delta_i, 1)$, is less than a fixed threshold value t, we can write for the i^{th} row

$$P_{i1} = \text{Prob}\{z < t\} = \int_{-\infty}^{t-\delta_i} \phi(v)\,dv = \Phi(t - \delta_i) \tag{11a}$$

Letting $\mu_i = t - \delta_i$, $\mu_i + 5$ is the probit transformation used in dose-response relationships (FINNEY, 1952) ; defining $\mu_i^* = \mu_i \pi/\sqrt{3}$, then

$$[1 + e^{-\mu_i^*}] \doteq \Phi(\mu_i^* \sqrt{3}/\pi) \tag{11b}$$

For $-5 < \mu_i < 5$, the difference between the left and right hand sides of (11b) does not exceed .022, being negligible from a practical point of view.

Suppose that a normal function is chosen to describe the probability of response. Let y_{i3} be the underlying variable, which under the conditions of the i^{th} row of the contingency table, is modeled as

$$y_{i3} = \mathbf{x}'_{i3}\boldsymbol{\beta}_3 + \mathbf{z}'_{i3}\mathbf{u}_3 + e_{i3} \tag{12a}$$

where $\mathbf{x}'_{i3}$ and $\mathbf{z}'_{i3}$ are known row vectors, $\boldsymbol{\beta}_3$ and $\mathbf{u}_3$ are unknown vectors, and e_{i3} is a residual. Likewise, the models for birth weight and pelvic opening are

$$y_{i1} = \mathbf{x}'_{i1}\boldsymbol{\beta}_1 + \mathbf{z}'_{i1}\mathbf{u}_1 + e_{i1} \tag{12b}$$

and

$$y_{i2} = \mathbf{x}'_{i2}\boldsymbol{\beta}_2 + \mathbf{z}'_{i2}\mathbf{u}_2 + e_{i2} \tag{12c}$$

Define μ_i in (9) as

$$\begin{aligned}\mu_i &= E(y_{i3}|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{u}_1, \mathbf{u}_2, \boldsymbol{\beta}_3, \mathbf{u}_3, y_{i1}, y_{i2}) \\ &= \mathbf{x}'_{i3}\boldsymbol{\beta}_3 + \mathbf{z}'_{i3}\mathbf{u}_3 + E(e_{i3}|e_{i1}, e_{i2})\end{aligned} \tag{13}$$

which holds if e_{i3} is correlated only with e_{i1} and e_{i2}. In a multivariate normal setting

$$E(e_{i3}|e_{i1}, e_{i2}) = [\rho_{31}\ \rho_{32}]\begin{bmatrix}1 & \rho_{12}\\ \rho_{12} & 1\end{bmatrix}^{-1}\begin{bmatrix}\sigma_{e_3}/\sigma_{e_1} & 0\\ 0 & \sigma_{e_3}/\sigma_{e_2}\end{bmatrix}\begin{bmatrix}e_{i1}\\ e_{i2}\end{bmatrix} \tag{14}$$

where the ρ_{ij}'s and the σ_{e_i}'s are residual correlations and residual standard deviations, respectively. Similarly

$$\begin{aligned}\mathrm{Var}(y_{i3}|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{u}_1, \mathbf{u}_2, \boldsymbol{\beta}_3, \mathbf{u}_3, y_{i1}, y_{i2}) &= \mathrm{Var}(e_{13}|e_{i1}, e_{i2}) \\ &= \sigma_3^2(1-\rho_{3.12}^2)\end{aligned} \tag{15}$$

where $\rho_{3.12}^2$ is the fraction of the residual variance of the underlying variable explained by a linear relationship with e_{i1} and e_{i2}. Since the unit of measurement in the conditional distribution of the underlying variate given $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{u}_1, \mathbf{u}_2, \boldsymbol{\beta}_3, \mathbf{u}_3, y_{i1}$ and y_{i2} is the standard deviation, then (14) can be written as

$$E(e_{i3}|e_{i1}, e_{i2}) = [\rho_{31}\rho_{32}]\begin{bmatrix}1 & \rho_{12}\\ \rho_{12} & 1\end{bmatrix}^{-1}\begin{bmatrix}1/\sigma_{e_1} & 0\\ 0 & 1/\sigma_{e_2}\end{bmatrix}\frac{1}{\sqrt{1-\rho_{3.12}^2}}\begin{bmatrix}e_{i1}\\ e_{i2}\end{bmatrix} \tag{16}$$

$$= b_1 e_{i1} + b_2 e_{i2} \tag{17}$$

Hence, (13) can be written in matrix notation as

$$\boldsymbol{\mu} = \mathbf{X}_3\boldsymbol{\beta}_3 + \mathbf{Z}_3\mathbf{u}_3 + b_1\mathbf{e}_1 + b_2\mathbf{e}_2 \tag{18}$$

$$= \mathbf{X}_3\boldsymbol{\beta}_3 + \mathbf{Z}_3\mathbf{u}_3 + \sum_{j=1}^{2} b_j(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta}_j - \mathbf{Z}_j\mathbf{u}_j) \tag{19}$$

where $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are known matrices arising from writing (12b) and (12c) as vectors. Now, suppose for simplicity that $\mathbf{X}_3$ is a matrix such that all factors and levels in $\mathbf{X}_1$ and $\mathbf{X}_2$ are represented in $\mathbf{X}_3$ and let $\mathbf{Z}_1 \equiv \mathbf{Z}_2 \equiv \mathbf{Z}_3$. Write

$$\mathbf{X}_1 = \mathbf{X}_3\mathbf{Q}_1; \qquad \mathbf{X}_2 = \mathbf{X}_3\mathbf{Q}_2$$

where $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are matrices of operators obtained by deleting columns of identity matrices of appropriate order. Thus, (19) can be written as

$$\begin{aligned}\boldsymbol{\mu} = \mathbf{X}_3(\boldsymbol{\beta}_3 - b_1\mathbf{Q}_1\boldsymbol{\beta}_1 - b_2\mathbf{Q}_2\boldsymbol{\beta}_2) + \mathbf{Z}_3(\mathbf{u}_3 - b_1\mathbf{u}_1 - b_2\mathbf{u}_2) \\ + b_1\mathbf{y}_1 + b_2\mathbf{y}_2\end{aligned} \tag{20}$$

Letting $\boldsymbol{\tau} = \boldsymbol{\beta}_3 - \sum_{j=1}^{2} b_j\mathbf{Q}_j\boldsymbol{\beta}_j$ and $\boldsymbol{\nu} = \mathbf{u}_3 - \sum_{j=1}^{2} b_j\mathbf{u}_j$, (20) can be expressed as

$$\boldsymbol{\mu} = \mathbf{X}_3\boldsymbol{\tau} + \mathbf{Z}_3\boldsymbol{\nu} + b_1\mathbf{y}_1 + b_2\mathbf{y}_2 \tag{21}$$

Note that if $b_1 = b_2 = 0$, then $\boldsymbol{\tau} = \boldsymbol{\beta}_3$, $\boldsymbol{\nu} = \mathbf{u}_3$, and (21) is equal to the expectation of (12a).

Given $\boldsymbol{\mu}$, the indicator variables $\mathbf{Y}$ are assumed to be conditionally independent, and the likelihood function is taken as product binomial so

$$\ln[f(\mathbf{Y}|\boldsymbol{\mu})] = \sum_{i=1}^{s}[n_{i1}\ln(P_{i1}) + (n_{i.} - n_{i1})\ln(1-P_{i1})] \tag{22}$$

Now

$$f(\mathbf{Y}|\boldsymbol{\mu}) \propto f(\mathbf{Y}|\mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\theta}^*) \tag{23}$$

where $\boldsymbol{\theta}^{*\prime} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{b}_1, \mathbf{b}_2]$. Also

$$f(\mathbf{Y}|\mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\theta}^*) \propto f(\mathbf{Y}|\mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\tau}, \mathbf{u}_1, \mathbf{u}_2, \boldsymbol{\nu}, b_1, b_2) \tag{24}.$$

Letting $\boldsymbol{\theta}' = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\tau}, \mathbf{u}_1, \mathbf{u}_2, \boldsymbol{\nu}, b_1, b_2]$, then from (23) and (24)

$$f(\mathbf{Y}|\boldsymbol{\mu}) \propto f(\mathbf{Y}|\mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\theta}^*) \propto f_6(\mathbf{Y}|\mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\theta}) \tag{25}.$$

B. *Conditional density of «risk» variables.*

The conditional density of $\mathbf{y}_1$ and $\mathbf{y}_2$ given $\boldsymbol{\theta}$ is assumed to be multivariate normal with location and dispersion following from (12b) and (12c)

$$\mathrm{E}\begin{bmatrix}\mathbf{y}_1|\boldsymbol{\theta} \\ \mathbf{y}_2|\boldsymbol{\theta}\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{Z}_1\mathbf{u}_1 \\ \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_2\mathbf{u}_2\end{bmatrix} \tag{26}$$

$$\mathrm{Var}\begin{bmatrix}\mathbf{y}_1|\boldsymbol{\theta} \\ \mathbf{y}_2|\boldsymbol{\theta}\end{bmatrix} = \begin{bmatrix}\mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22}\end{bmatrix} \tag{27}$$

where (27) is a non-singular known covariance matrix. Letting $\mathbf{R}^{11}$, $\mathbf{R}^{12}$, $\mathbf{R}^{21}$ and $\mathbf{R}^{22}$ be respective partitions of the inverse of (27), one can write

$$\ln[f_5(\mathbf{y}_1, \mathbf{y}_2|\boldsymbol{\theta})] = -\frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_i - \mathbf{Z}_i\mathbf{u}_i)\mathbf{R}^{ij}(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta}_j - \mathbf{Z}_j\mathbf{u}_j) + \text{const.} \tag{28}$$

C. *Prior density.*

In this paper we assume that the residual covariance matrix

$$\mathrm{Var}\begin{bmatrix}\mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3\end{bmatrix} = \begin{bmatrix}\mathbf{R}_{11} & \mathbf{R}_{12} & \mathbf{R}_{13} \\ \mathbf{R}_{21} & \mathbf{R}_{22} & \mathbf{R}_{23} \\ \mathbf{R}_{31} & \mathbf{R}_{32} & \mathbf{R}_{33}\end{bmatrix}$$

is known. From (16) and (17), this implies that b_1 and b_2 are also known. Therefore,

$$f(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\tau}, \mathbf{u}_1, \mathbf{u}_2, \boldsymbol{\nu}, b_1, b_2) \propto f(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\tau}, \mathbf{u}_1, \mathbf{u}_2, \boldsymbol{\nu}) \tag{28}$$

and the vector of unknowns becomes $\boldsymbol{\theta}' = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\tau}, \mathbf{u}_1, \mathbf{u}_2, \boldsymbol{\nu}]$

Let $\boldsymbol{\theta}' = [\boldsymbol{\beta}', \mathbf{u}']$ where $\boldsymbol{\beta}' = [\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\tau}']$ and $\mathbf{u}' = [\mathbf{u}_1', \mathbf{u}_2', \boldsymbol{\nu}']$. *A priori* $\boldsymbol{\theta}$ follows multivariate normal distribution

$$\boldsymbol{\theta} = \begin{bmatrix}\boldsymbol{\beta} \\ \mathbf{u}\end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix}\boldsymbol{\alpha} \\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Gamma} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_c\end{bmatrix}\right) \tag{29}$$

where

$$\mathrm{Var}(\mathbf{u}) = \mathrm{Var}\begin{bmatrix}\mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 - b_1\mathbf{u}_1 - b_2\mathbf{u}_2\end{bmatrix} = \mathbf{G}_c \tag{30}$$

Further

$$\mathbf{G}_c = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ -b_1\mathbf{I} & -b_2\mathbf{I} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} & \mathbf{G}_{13} \\ \mathbf{G}_{21} & \mathbf{G}_{22} & \mathbf{G}_{23} \\ \mathbf{G}_{31} & \mathbf{G}_{32} & \mathbf{G}_{33} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} & -b_1\mathbf{I} \\ \mathbf{0} & \mathbf{I} & -b_2\mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{31}$$

with $\mathrm{Cov}(\mathbf{u}_i, \mathbf{u}_j') = \mathbf{G}_{ij}$ (i, j = 1, ..., 3). Note that $\mathbf{G}_c$ depends on b_1 and b_2; when $b_1 = b_2 = 0$, it follows from (30) that $\mathbf{G}_c = \{\mathbf{G}_{ij}\}$. Now

$$\ln[f_1(\theta)] = -\frac{1}{2}\{(\boldsymbol{\beta}-\boldsymbol{\alpha})'\Gamma^{-1}(\boldsymbol{\beta}-\boldsymbol{\alpha}) + \mathbf{u}'\mathbf{G}_c^{-1}\mathbf{u}\} + \text{const.} \tag{32}$$

where $\mathbf{G}_c^{-1} = \{\mathbf{G}_c^{ij}\}$ (i, j = 1, ..., 3). Prior knowledge about $\boldsymbol{\beta}$ is assumed to be vague so $\Gamma \longrightarrow \infty$ and $\Gamma^{-1} \longrightarrow 0$. Therefore

$$\ln[f_1(\theta)] = -\frac{1}{2}\{\mathbf{u}'\mathbf{G}_c^{-1}\mathbf{u}\} + \text{const.} \tag{33}.$$

IV. Estimation

The terms of the log-posterior density in (8) are given in equations (22), (28) and (33). To obtain the mode of the posterior density, the derivatives of (8) with respect to $\boldsymbol{\theta}$ are equated to zero. The resulting system of equations is not linear in $\boldsymbol{\theta}$ and an iterative solution is required. Letting $L(\boldsymbol{\theta})$ be the log of the posterior density, the Newton-Raphson algorithm (DAHLQUIST & BJÖRCK, 1974) consists of iterating with

$$\hat{\boldsymbol{\theta}}^{[i]} = \hat{\boldsymbol{\theta}}^{[i-1]} - \left[\left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]\right]^{-1}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{[i-1]}} \left[\left[\frac{\partial L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{[i-1]}} \tag{34}.$$

Note that the inverse of the matrix of second partial derivatives exists as $\boldsymbol{\beta}$ can be uniquely defined, e.g., with $\mathbf{X}_i$ having full-column rank, i = 1, ...3. It is convenient to write (34) as

$$\left[\left[-\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{[i-1]}} (\hat{\boldsymbol{\theta}}^{[i]} - \hat{\boldsymbol{\theta}}^{[i-1]}) = \left[\left[\frac{\partial L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{[i-1]}} \tag{35}.$$

A. *First derivatives.*

Differentiating (8) with respect to the elements of $\boldsymbol{\theta}$ yields

$$\frac{\partial L(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}_i} = \mathbf{X}_i' \sum_{j=1}^{2} \mathbf{R}^{ij}(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta}_j - \mathbf{Z}_j\mathbf{u}_j),\ i = 1,2 \tag{36}$$

and

$$\frac{\partial L(\boldsymbol{\theta})}{\partial\mathbf{u}_i} = \mathbf{Z}_i' \sum_{j=1}^{2} \mathbf{R}^{ij}(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta}_j - \mathbf{Z}_j\mathbf{u}_j) - \sum_{j=1}^{2} \mathbf{G}_c^{ij}\mathbf{u}_j - \mathbf{G}_c^{i3}\boldsymbol{\nu} \tag{37}.$$

The derivatives of $L(\boldsymbol{\theta})$ with respect to $\boldsymbol{\tau}$ and $\boldsymbol{\nu}$ are slightly different

$$\frac{\partial L(\boldsymbol{\theta})}{\partial\boldsymbol{\tau}} = \sum_{i=1}^{s} \left[n_{i1}\frac{\phi(\mu_i)}{P_{i1}} - (n_{i\cdot} - n_{i1})\frac{\phi(\mu_i)}{1-P_{i1}}\right]\mathbf{x}_{i3} \tag{38}$$

where $\mathbf{x}'_{i3}$ is the i^{th} row of $\mathbf{X}_3$, and

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\nu}} = \sum_{i=1}^{s} \left[n_{i1} \frac{\phi(\mu_i)}{P_{i1}} - (n_{i.} - n_{i1}) \frac{\phi(\mu_i)}{1 - P_{i1}} \right] \mathbf{z}_{i3} - \sum_{j=1}^{2} \mathbf{G}_c^{3j} \mathbf{u}_j - \mathbf{G}_c^{33} \boldsymbol{\nu}. \tag{39}$$

Now, let $\mathbf{v}$ be a sx1 vector with elements

$$v_j = -\{n_{j1} i_{j1} + (n_{j.} - n_{j1}) i_{j2}\}, \quad j = 1, \ldots, s$$

where $i_{j1} = -\phi(\mu_j)/P_{j1}$ and $i_{j2} = \phi(\mu_j)/(1 - P_{j1})$, and note that v_j is the opposite of the sum of normal scores for the j^{th} row. Then

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}} = \mathbf{X}'_3 \mathbf{v} \tag{40}$$

and

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\nu}} = \mathbf{Z}'_3 \mathbf{v} - \sum_{j=1}^{2} \mathbf{G}_c^{3j} \mathbf{u}_j - \mathbf{G}_c^{33} \boldsymbol{\nu}. \tag{41}$$

B. *Second derivatives*

The symmetric matrix of second partial derivatives can be deduced from equations (36) through (41). Explicitly

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\beta}'_i} = -\mathbf{X}'_i \mathbf{R}^{ii} \mathbf{X}_i; \qquad i = 1,2 \tag{42a}$$

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}'_2} = -\mathbf{X}'_1 \mathbf{R}^{12} \mathbf{X}_2 \tag{42b}$$

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_i \partial \mathbf{u}'_j} = -\mathbf{X}'_i \mathbf{R}^{ij} \mathbf{Z}_j; \qquad \begin{matrix} i = 1,2 \\ j = 1,2 \end{matrix} \tag{42c}$$

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\tau}'} = \mathbf{0}\ ; \qquad i = 1,2 \tag{42d}$$

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\nu}'} = \mathbf{0}\ ; \qquad i = 1,2 \tag{42e}$$

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \mathbf{u}_i \partial \mathbf{u}'_j} = -\mathbf{Z}'_i \mathbf{R}^{ij} \mathbf{Z}_j - \mathbf{G}_c^{ij}; \qquad \begin{matrix} i = 1,2 \\ j = 1,2 \end{matrix} \tag{42f}$$

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \mathbf{u}_i \partial \boldsymbol{\tau}'} = \mathbf{0}\ ; \qquad i = 1,2 \tag{42g}$$

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \mathbf{u}_i \partial \boldsymbol{\nu}'} = -\mathbf{G}^{i3}; \qquad i = 1,2 \tag{42h}$$

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'} = -\mathbf{X}'_3 \mathbf{W} \mathbf{X}_3 \tag{42i}$$

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\nu}'} = -\mathbf{X}'_3 \mathbf{W} \mathbf{Z}_3; \tag{42j}$$

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}'} = -\mathbf{Z}'_3 \mathbf{W} \mathbf{Z}_3 - \mathbf{G}_c^{33}. \tag{42k}$$

In (42 *i*) through (42 *k*), $\mathbf{W}$ is an sxs diagonal matrix with elements

$$w_{jj} = n_{j1} i_{j1} (i_{j1} - \mu_j) + (n_{j.} - n_{j1}) i_{j2} (i_{j2} - \mu_j)$$
$$= \mu_j v_j + n_{j1} i_{j1}^2 + (n_{j.} - n_{j1}) i_{j2}^2, \; j = 1, \ldots, s. \tag{43}$$

Note that
$$E(w_{jj} | \mu_j) = n_{j1} i_{j1}^2 + (n_{j.} - n_{j1}) i_{j2}^2 \tag{44}$$

indicating that calculations are somewhat simpler if «scoring» is used instead of Newton-Raphson.

C. *Equations*

Using the first and second derivatives in (36-41) and (42 *a*-42 *k*), respectively, equations (35) can be written after algebra as (45).

In (45), $\hat{\boldsymbol{\beta}}_1^{[i]}$, $\hat{\boldsymbol{\beta}}_2^{[i]}$, $\hat{\mathbf{u}}_1^{[i]}$ and $\hat{\mathbf{u}}_2^{[i]}$ are solutions at the [ith] iterate while the $\boldsymbol{\Delta}$'s are corrections at the [ith] iterate pertaining to the parameters affecting the probability of response, e.g., $\boldsymbol{\Delta}_\tau^{[i]} = \hat{\boldsymbol{\tau}}^{[i]} - \hat{\boldsymbol{\tau}}^{[i-1]}$. Iteration proceeds by first taking a guess for $\boldsymbol{\tau}$ and $\boldsymbol{\nu}$, calculating $\mathbf{W}^{[0]}$ and $\mathbf{v}^{[0]}$, amending the right hand-sides and then solving for the unknowns. The cycle is repeated until the solutions stabilize. Equations (45) can also be written as in (46). The similarity between (46) and the «mixed model equations» (HENDERSON, 1973) should be noted. The coefficient matrix and the «working» vector $\mathbf{y}_3^{[i-1]}$ change in every iteration; note that $\mathbf{y}_3^{[i-1]} = \mathbf{X}_3 \boldsymbol{\tau}^{[i-1]} + \mathbf{Z}_3 \boldsymbol{\nu}^{[i-1]} + (\mathbf{W}^{[i-1]})^{-1} \mathbf{v}^{[i-1]}$.

D. *Solving the equations*

In animal breeding practice, solving (45) or (46) poses a formidable numerical problem. The order of the coefficient matrix can be in the tens of thousands, and this difficulty arises in every iterate. As $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\mathbf{u}_1$ and $\mathbf{u}_2$ are «nuisance» variables in this problem, the first step is to eliminate them from the system, if this is feasible. The order of the remaining equations is still very large in most animal breeding problems so direct inversion is not possible. At the ith iterate, the remaining equations can be written as

$$\mathbf{P}^{[i-1]} \boldsymbol{\gamma}^{[i]} = \mathbf{I}^{[i-1]} \tag{47}.$$

Next, decompose $\mathbf{P}^{[i-1]}$ as the sum of three matrices $\mathbf{L}^{[i-1]}$, $\mathbf{D}^{[i-1]}$, $\mathbf{U}^{[i-1]}$, which are lower triangular, diagonal and upper triangular, respectively. Therefore

$$\boldsymbol{\gamma}^{[i]} = \{\mathbf{D}^{[i-1]}\}^{-1} \{\mathbf{I}^{[i-1]} - \mathbf{L}^{[i-1]} \boldsymbol{\gamma}^{[i]} - \mathbf{U}^{[i-1]} \boldsymbol{\gamma}^{[i]}\}.$$

Now, for *each* iterate *i*, sub-iterate with

$$\boldsymbol{\gamma}^{[i,\, j+1]} = \{\mathbf{D}^{[i-1]}\}^{-1} \{\mathbf{I}^{[i-1]} - \mathbf{L}^{[i-1]} \boldsymbol{\gamma}^{[i,\, j+1]} - \mathbf{U}^{[i-1]} \boldsymbol{\gamma}^{[i,\, j]}\} \tag{48}$$

for j = 0, 1, ...; iteration can start with $\boldsymbol{\gamma}^{[i,\, 0]} = \mathbf{0}$. As this is a «nested» Gauss-Seidel iteration, with $\mathbf{P}^{[i-1]}$ symmetric and positive definite

$$\lim_{j \to \infty} \boldsymbol{\gamma}^{[i,\, j]} = \boldsymbol{\gamma}^{[i]} \tag{49}$$

(VAN NORTON, 1960). Then, one needs to return to (47) and to the back solution, and work with (48). The cycle finishes when the solutions $\boldsymbol{\gamma}$ stabilize.

$$
\begin{bmatrix}
\mathbf{X}_1'\mathbf{R}^{11}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}^{12}\mathbf{X}_2 & \mathbf{X}_1'\mathbf{R}^{11}\mathbf{Z}_1 & \mathbf{X}_1'\mathbf{R}^{12}\mathbf{Z}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{X}_2'\mathbf{R}^{21}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{R}^{22}\mathbf{X}_2 & \mathbf{X}_2'\mathbf{R}^{21}\mathbf{Z}_1 & \mathbf{X}_2'\mathbf{R}^{22}\mathbf{Z}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{Z}_1'\mathbf{R}^{11}\mathbf{X}_1 & \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{X}_2 & \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{Z}_1+\mathbf{G}_c^{11} & \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{Z}_2+\mathbf{G}_c^{12} & \mathbf{0} & \mathbf{G}_c^{13} & \mathbf{0} \\
\mathbf{Z}_2'\mathbf{R}^{21}\mathbf{X}_1 & \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{X}_2 & \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{Z}_1+\mathbf{G}_c^{21} & \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{Z}_2+\mathbf{G}_c^{22} & \mathbf{0} & \mathbf{G}_c^{23} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{X}_3'\mathbf{W}^{[i-1]}\mathbf{X}_3 & \mathbf{X}_3'\mathbf{W}^{[i-1]}\mathbf{Z}_3 & \mathbf{X}_3'\mathbf{W}^{[i-1]}\mathbf{Z}_3+\mathbf{G}_c^{33}
\end{bmatrix}
\begin{bmatrix}
\widehat{\boldsymbol{\beta}}_1^{[i]} \\ \widehat{\boldsymbol{\beta}}_2^{[i]} \\ \widehat{\mathbf{u}}_1^{[i]} \\ \widehat{\mathbf{u}}_2^{[i]} \\ \boldsymbol{\Delta}_\nu^{[i]}
\end{bmatrix}
$$

$$
= \begin{bmatrix}
\mathbf{X}_1'(\mathbf{R}^{11}\mathbf{y}_1+\mathbf{R}^{12}\mathbf{y}_2) \\
\mathbf{X}_2'(\mathbf{R}^{21}\mathbf{y}_1+\mathbf{R}^{22}\mathbf{y}_2) \\
\mathbf{Z}_1'(\mathbf{R}^{11}\mathbf{y}_1+\mathbf{R}^{12}\mathbf{y}_2) \\
\mathbf{Z}_2'(\mathbf{R}^{21}\mathbf{y}_1+\mathbf{R}^{22}\mathbf{y}_2) \\
\mathbf{X}_3'\boldsymbol{\nu}^{[i-1]} \\
\mathbf{Z}_3'\boldsymbol{\nu}^{[i-1]}
\end{bmatrix}
- \begin{bmatrix}
\mathbf{0} \\ \mathbf{0} \\ \mathbf{G}_c^{13}\boldsymbol{\nu}^{[i-1]} \\ \mathbf{G}_c^{23}\boldsymbol{\nu}^{[i-1]} \\ \mathbf{0} \\ \mathbf{G}_c^{33}\boldsymbol{\nu}^{[i-1]}
\end{bmatrix} \tag{45}
$$

$$
\begin{bmatrix}
\mathbf{X}_1'\mathbf{R}^{11}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}^{12}\mathbf{X}_2 & \mathbf{X}_1'\mathbf{R}^{11}\mathbf{Z}_1 & \mathbf{X}_1'\mathbf{R}^{12}\mathbf{Z}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{X}_2'\mathbf{R}^{21}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{R}^{22}\mathbf{X}_2 & \mathbf{X}_2'\mathbf{R}^{21}\mathbf{Z}_1 & \mathbf{X}_2'\mathbf{R}^{22}\mathbf{Z}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{Z}_1'\mathbf{R}^{11}\mathbf{X}_1 & \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{X}_2 & \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{Z}_1+\mathbf{G}_c^{11} & \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{Z}_2+\mathbf{G}_c^{12} & \mathbf{0} & \mathbf{G}_c^{13} & \mathbf{0} \\
\mathbf{Z}_2'\mathbf{R}^{21}\mathbf{X}_1 & \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{X}_2 & \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{Z}_1+\mathbf{G}_c^{21} & \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{Z}_2+\mathbf{G}_c^{22} & \mathbf{0} & \mathbf{G}_c^{23} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{X}_3'\mathbf{W}^{[i-1]}\mathbf{X}_3 & \mathbf{X}_3'\mathbf{W}^{[i-1]}\mathbf{Z}_3 & \mathbf{X}_3'\mathbf{W}^{[i-1]}\mathbf{Z}_3+\mathbf{G}_c^{33}
\end{bmatrix}
\begin{bmatrix}
\widehat{\boldsymbol{\beta}}_1^{[i]} \\ \widehat{\boldsymbol{\beta}}_2^{[i]} \\ \widehat{\mathbf{u}}_1^{[i]} \\ \widehat{\mathbf{u}}_2^{[i]} \\ \boldsymbol{\tau}^{[i]} \\ \boldsymbol{\nu}^{[i]}
\end{bmatrix}
$$

$$
= \begin{bmatrix}
\mathbf{X}_1'(\mathbf{R}^{11}\mathbf{y}_1+\mathbf{R}^{12}\mathbf{y}_2) \\
\mathbf{X}_2'(\mathbf{R}^{21}\mathbf{y}_1+\mathbf{R}^{22}\mathbf{y}_2) \\
\mathbf{Z}_1'(\mathbf{R}^{11}\mathbf{y}_1+\mathbf{R}^{12}\mathbf{y}_2) \\
\mathbf{Z}_2'(\mathbf{R}^{21}\mathbf{y}_1+\mathbf{R}^{22}\mathbf{y}_2) \\
\mathbf{X}_3'\mathbf{W}^{[i-1]}\mathbf{y}_3^{[i-1]} \\
\mathbf{Z}_3'\mathbf{W}^{[i-1]}\mathbf{y}_3^{[i-1]}
\end{bmatrix} \tag{46}
$$

Another possibility would be to carry out nested iterations with the conjugate gradient method (BECKMAN, 1960). In the context of (47) the method involves :

a) Set

$$\mathbf{P}^{[0]} = \mathbf{r}^{[0]} = \mathbf{l}^{[i-1]} - \mathbf{P}^{[i-1]}\boldsymbol{\gamma}^{[i.\,0]};\ i = 1,\ 2,\ \ldots$$

where $\boldsymbol{\gamma}^{[i.\,0]}$ is a guess, e.g., $\boldsymbol{\gamma}^{[i.\,0]} = 0$.

b) Calculate successively

$$\begin{aligned}
&[b.1] \quad \alpha^{[j]} &&= \mathbf{p}'^{[j]}\mathbf{r}^{[j]} / \left\{\mathbf{p}'^{[j]}\mathbf{P}^{[i-1]}\mathbf{p}^{[j]}\right\} \\
&[b.2] \quad \boldsymbol{\gamma}^{[i.\,j+1]} &&= \boldsymbol{\gamma}^{[i.\,j]} + \alpha^{[j]}\mathbf{p}^{[j]} \\
&[b.3] \quad \mathbf{r}^{[j+1]} &&= \mathbf{r}^{[j]} - \alpha^{[j]}\mathbf{p}^{[i-1]}\mathbf{p}^{[j]} \\
&[b.4] \quad \lambda^{[j]} &&= -\left\{\mathbf{r}'^{[j+1]}\mathbf{P}^{[i-1]}\mathbf{p}^{[j]}\right\} / \left\{\mathbf{p}'^{[j]}\mathbf{P}^{[i-1]}\mathbf{p}^{[j]}\right\} \\
&[b.5] \quad \mathbf{p}^{[j+1]} &&= \mathbf{r}^{[j+1]} + \lambda^{[j]}\mathbf{p}^{[j]}
\end{aligned}$$

for $j = 0,\ 1,\ \ldots$, until $\gamma^{[i.\,j]}$ stabilizes. When this occurs, $\mathbf{P}^{[i-1]}$ and $\mathbf{l}^{[i-1]}$ in (47) are amended, and the cycle with a new index for i is started from (a). The whole process stops when $\gamma^{[i]}$ does not change between the [i] and [i + 1] « main » rounds. While the number of operations per iterate is higher than with Gauss-Seidel (BECKMAN, 1960), the method is known to converge faster when $\mathbf{P}^{[i-1]}$ in (47) is symmetric and positive definite (personal communication, SAMEH, 1981).

V. Approximate posterior inference and model fit

As discussed by LINDLEY & SMITH (1972) in the context of linear models, the procedure does not provide standard errors a posteriori. LEONARD (1972), however, has pointed out that an approximation of the posterior density by a multivariate normal is « fairly accurate » in most regions of the space of $\boldsymbol{\theta}$, provided that none of the n_{i1} or $n_{i.} - n_{i1}$ are small. If this approximation can be justified, given any linear function of $\boldsymbol{\theta}$, say $\mathbf{t}'\boldsymbol{\theta}$, one can write, given the model

$$\mathrm{E}[\mathbf{t}'\boldsymbol{\theta} \,|\, \mathbf{Y},\ \mathbf{y}_1,\ \mathbf{y}_2] = \mathbf{t}'\hat{\boldsymbol{\theta}} \tag{50a}$$

$$\mathrm{Var}[\mathbf{t}'\boldsymbol{\theta} \,|\, \mathbf{Y},\ \mathbf{y}_1,\ \mathbf{y}_2] = \mathbf{t}'\mathbf{C}\mathbf{t} \tag{50b}$$

where $\hat{\boldsymbol{\theta}}$ is the posterior mode and $\mathbf{C}$ is the inverse of the coefficient matrix in (46); note that $\mathbf{C}$ depends on the data through the matrix $\mathbf{W}$. Further

$$[\mathbf{t}'\boldsymbol{\theta} - \mathbf{t}'\hat{\boldsymbol{\theta}})/(\mathbf{t}'\mathbf{C}\mathbf{t})^{\frac{1}{2}} \sim \mathrm{N}(0,\ 1) \tag{50c}$$

thus permitting probability statements about $\mathbf{t}'\boldsymbol{\theta}$. In many instances it will be impossible to calculate $\mathbf{C}$ on computational grounds.

The probability of response for each of the rows in the contingency table can be estimated from (9) with μ evaluated at $\hat{\mu}$. Approximate standard errors of the estimates of response probabilities can be obtained from large sample theory. However, caution should be exercised as an approximation to an approximation is involved.

When cell counts are large, e.g., n_{i1} and $n_{i.} - n_{i1} > 5$, the statistic

$$\chi^2 = \sum_{i=1}^{s} \frac{(n_{i1} - n_i \hat{P}_{i1})^2}{n_{i.}\hat{P}_{i1}(1 - \hat{P}_{i1})} \tag{51}$$

can be referred to a chi-square distribution with s-rank $(\mathbf{X}_3)$ degrees of freedom. Lack of fit may result from inadequate model specification in which case alternative models should be entertained.

VI. Unknown variance-covariance structure

The matrices $\mathbf{R}_{ij}(i, j = 1, \ldots, 3)$ and $\mathbf{G}_c$ are assumed known so that they are treated as nuisance arrays in (8) and (46). In animal breeding practice there are generally « good » estimates of these matrices so they could be used in (45) or (46) to proceed with the method, in the same way as in linear methodology (HENDERSON, 1973). The effect of replacing $\mathbf{R}$ and $\mathbf{G}_c$ matrices by estimates on the posterior distribution of $\boldsymbol{\theta}$ is not known, and should be studied by Monte-Carlo methods.

If the analysis were to proceed in an entirely Bayesian context, prior distributions would need to be specified for the elements of these matrices. This is not addressed in the present paper as it does not appear clear what densities should be considered for the distribution of covariance components. For a discussion of Bayes estimation of variance components, see HILL (1965), TIAO & TAN (1965), TIAO & BOX (1967), LINDLEY & SMITH (1972) and HARVILLE (1977). LEONARD (1972) considered estimation of variance components with binomial data for a one-way model.

Equations (46) suggest methods for estimating variance and covariance components in this quantitative-categorical setting. Write

a) $\mathbf{W}^{[i]} = (\mathbf{A}^{[i]})^2 = \left(\left\{\sqrt{w_{jj}^{[i]}}\right\}\right)^2$, $j = 1, \ldots, s$
b) $\mathbf{X}_3^{[i]} = \mathbf{A}^{[i]}\mathbf{X}_3$
c) $\mathbf{Z}_3^{[i]} = \mathbf{A}^{[i]}\mathbf{Z}_3$

and

d) $\mathbf{q}_3^{[i]} = \mathbf{A}^{[i]}\mathbf{y}_3^{[i]}$

Equations (46) can then be written as (52) below.

$$\begin{bmatrix} \mathbf{X}_1'\mathbf{R}^{11}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}^{12}\mathbf{X}_2 & \mathbf{X}_1'\mathbf{R}^{11}\mathbf{Z}_1 & \mathbf{X}_1'\mathbf{R}^{12}\mathbf{Z}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{X}_2'\mathbf{R}^{21}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{R}^{22}\mathbf{X}_2 & \mathbf{X}_2'\mathbf{R}^{21}\mathbf{Z}_1 & \mathbf{X}_2'\mathbf{R}^{22}\mathbf{Z}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{X}_1 & \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{X}_2 & \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{Z}_1 + \mathbf{G}_c^{11} & \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{Z}_2 + \mathbf{G}_c^{12} & \mathbf{0} & \mathbf{G}_c^{13} \\ \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{X}_1 & \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{X}_2 & \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{Z}_1 + \mathbf{G}_c^{21} & \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{Z}_2 + \mathbf{G}_c^{22} & \mathbf{0} & \mathbf{G}_c^{23} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{X}_3^{[i]}\mathbf{X}_3^{[i]} & \mathbf{X}_3^{[i]}\mathbf{Z}_3^{[i]} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}_c^{31} & \mathbf{G}_c^{32} & \mathbf{Z}_3^{[i]}\mathbf{X}_3^{[i]} & \mathbf{Z}_3^{[i]}\mathbf{Z}_3^{[i]} + \mathbf{G}_c^{33} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_1^{[i+1]} \\ \hat{\boldsymbol{\beta}}_2^{[i+1]} \\ \hat{\mathbf{u}}_1^{[i+1]} \\ \hat{\mathbf{u}}_2^{[i+1]} \\ \hat{\boldsymbol{\tau}}^{[i+1]} \\ \hat{\boldsymbol{\nu}}^{[i+1]} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{X}_1'(\mathbf{R}^{11}\mathbf{y}_1 + \mathbf{R}^{12}\mathbf{y}_2) \\ \mathbf{X}_2'(\mathbf{R}^{21}\mathbf{y}_1 + \mathbf{R}^{22}\mathbf{y}_2) \\ \mathbf{Z}_1'(\mathbf{R}^{11}\mathbf{y}_1 + \mathbf{R}^{12}\mathbf{y}_2) \\ \mathbf{Z}_2'(\mathbf{R}^{21}\mathbf{y}_1 + \mathbf{R}^{22}\mathbf{y}_2) \\ \mathbf{X}_3'^{[i]}\mathbf{q}^{[i]} \\ \mathbf{Z}_3'^{[i]}\mathbf{q}^{[i]} \end{bmatrix} \qquad (52)$$

The above equations suggest at each iterate the multivariate linear model

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{q}^{[i]} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_3^{[i]} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1^{[i+1]} \\ \boldsymbol{\beta}_2^{[i+1]} \\ \boldsymbol{\tau}^{[i+1]} \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}_3^{[i]} \end{bmatrix} \begin{bmatrix} \mathbf{u}^{[i+1]} \\ \mathbf{u}_2^{[i+1]} \\ \boldsymbol{\nu}^{[i+1]} \end{bmatrix} + \begin{bmatrix} \mathbf{E}_1^{[i]} \\ \mathbf{E}_2^{[i]} \\ \mathbf{E}_3^{[i]} \end{bmatrix}$$

$$i = 0, 1 \ldots \qquad (53)$$

with $\boldsymbol{\beta}_1^{[i+1]}$, $\boldsymbol{\beta}_2^{[i+1]}$ and $\boldsymbol{\tau}^{[i+1]}$ « fixed » and $\mathbf{u}_1^{[i+1]}$, $\mathbf{u}_2^{[i+1]}$, $\boldsymbol{\nu}^{[i+1]}$ and the $\mathbf{E}$'s random, with covariance matrix

$$\text{Var}\begin{bmatrix}\mathbf{u}_1^{[i+1]}\\ \mathbf{u}_2^{[i+1]}\\ \boldsymbol{\nu}^{[i+1]}\\ \mathbf{E}_1^{[i]}\\ \mathbf{E}_2^{[i]}\\ \mathbf{E}_3^{[i]}\end{bmatrix} = \begin{bmatrix}\mathbf{G}_{11.c} & \mathbf{G}_{12.c} & \mathbf{G}_{13.c} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{G}_{21.c} & \mathbf{G}_{22.c} & \mathbf{G}_{23.c} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{G}_{31.c} & \mathbf{G}_{32.c} & \mathbf{G}_{33.c} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_{11} & \mathbf{R}_{12} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_{21} & \mathbf{R}_{22} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}\end{bmatrix}$$

$$i = 0, 1 \ldots \quad (54)$$

holding at every iterate. Note that the residual variance of $\mathbf{q}^{[i]}$ is unity so this part of the covariance structure does not need to be estimated. Provided that ρ_{31} and ρ_{32} are known, the method can be used to estimate the additive genetic covariance matrix between the quantitative traits and the hypothetical underlying variate with binary expression.

Expressions in (53) and (54) suggest that some of the methods for estimating variance and covariance components in linear models could be used to estimate the covariance structure in (54). One possibility would be to mimic the computations used in estimation via restricted maximum likelihood (SCHAEFFER *et al.*, 1978) for multivariate normal data. As computational feasibility is of paramount importance, a multivariate extension of Henderson's « simple » method (HENDERSON, 1980) could be useful here. However, this method does not preclude negative estimates of variance components. Estimation of genetic parameters in non-linear models is an open area of potential importance.

VII. Numerical application

Data were obtained from 47 *Blonde d'Aquitaine* heifers mated to the same bull and assembled to calve in the Casteljaloux Station, France. Each calving record included information on the following: region of origin and sire of the heifer, pelvic opening and season of calving, sex and birth weight of the calf, and calving difficulty score (1: normal birth, 2: slight assistance, 3: assisted, 4: mechanical aid, and 5: cesarean). For the purpose of the analysis, twin calves were excluded and calving difficulty was recoded as: a) « Easy » (scores 1, 2 and 3) or b) « Difficult » (scores 4 and 5). The data are presented in Table 1. As shown in Table 2, 23.4 % of the calvings were « difficult » and there were marked differences in the incidence of difficult calvings between sexes and maternal grandsires.

A. *Models*

Birth weight was modeled as

$$Y_{ijklm} = D_i + T_j + L_k + S_l + e_{ijklm} \quad (55\,a)$$

where D_i is the effect of the i^{th} region of origin of the heifer ($i = 1,2$), T_j is the effect of the j^{th} season of calving ($j = 1,2$), L_k is the effect of the k^{th} sex of calf ($k = 1$: male, $2 =$ female), S_l is the effect of the l^{th} sire of the heifer ($l = 1, \ldots, 6$), and e_{ijklm} is a residual. The vectors $\boldsymbol{\beta}_l$ and $\mathbf{u}_l$ were defined as

$$\boldsymbol{\beta}_l' = [D_1 + T_2 + L_2,\ D_2 + T_2 + L_2,\ T_1 - T_2,\ L_1 - L_2] \quad (55\,b)$$

and

$$\mathbf{u}_l' = [S_1, S_2, S_3, S_4, S_5, S_6]. \quad (55\,c)$$

TABLE 1

Calving difficulty records for Blonde d'Aquitaine *heifers at Casteljaloux France.*

Données de vêlage des génisses de race Blonde d'Aquitaine *contrôlées à Casteljaloux, France.*

Heifer Origin	Sire of heifer	Calving season	Pelvic opening (cm^2)	Sex[a] of calf	Calf's birth weight (kg)	Calving[b] difficulty	Heifer Origin	Sire of heifer	Calving season	Pelvic opening (cm^2)	Sex[a] of calf	Calf's birth weight (kg)	Calving[b] difficulty
1	1	1	328.0	M	41.0	E	1	4	2	328.0	M	47.0	D
1	1	1	304.5	M	37.5	E	1	4	2	344.0	F	51.0	D
1	1	1	354.8	F	41.5	E	1	4	2	243.0	F	39.0	E
1	1	2	374.0	F	40.0	E	2	4	1	333.3	M	44.5	E
1	1	2	285.0	F	43.0	E	1	5	1	346.5	M	40.5	E
1	1	2	310.0	F	42.0	E	1	5	1	304.5	F	43.5	E
1	1	2	270.0	F	35.0	E	1	5	2	396.0	M	42.5	E
2	1	1	336.0	F	46.0	E	1	5	2	285.0	M	48.8	D
2	1	1	374.0	F	40.5	E	1	5	2	328.0	M	38.5	E
2	1	2	346.5	F	39.0	E	1	5	2	346.5	M	52.0	E
1	2	1	346.5	M	41.4	E	1	5	2	344.0	F	48.0	E
1	2	1	333.3	M	43.0	D	2	5	1	357.0	F	41.0	E
1	2	2	346.5	F	34.0	E	2	5	1	357.0	M	50.5	D
1	2	2	307.5	M	47.0	D	2	5	2	317.8	M	43.7	D
1	2	2	315.0	M	42.0	E	2	5	2	346.5	M	51.0	D
2	2	2	300.0	M	44.5	E	1	6	1	290.0	F	51.6	D
2	2	2	273.0	M	49.0	E	1	6	1	260.0	M	45.3	D
1	3	1	341.0	M	41.6	E	1	6	1	357.0	F	36.5	E
2	3	1	300.0	M	36.0	E	1	6	2	354.8	M	50.5	E
2	3	1	328.0	F	42.7	E	1	6	2	273.0	M	46.0	D
2	3	2	266.0	F	32.5	E	1	6	2	315.0	M	45.0	E
2	3	2	302.3	F	44.4	E	1	6	2	256.5	F	36.0	E
2	3	2	320.0	M	46.0	E	2	6	1	317.8	F	43.5	E
							2	6	1	290.0	F	36.5	E

a) M: male; F: female.
b) D: difficult; E: easy.

TABLE 2

Marginal mean (or frequency) of calving variables by level of factors considered. Moyenne (ou fréquence) marginale des variables de vêlage par niveau de facteur.

Factor		N°	Variable[a]		
			BW	PO	CD
Heifer origin	1	30	43.02	319.60	0.267
	2	17	43.02	321.48	0.177
Calving season	1	20	42.23	327.96	0.200
	2	27	43.61	314.59	0.259
Sex of calf	M	25	44.59	-	0.360
	F	22	41.24	-	0.091
Sire of heifer	1	10	40.55	328.28	0.000
	2	7	42.99	317.40	0.286
	3	6	40.53	309.55	0.000
	4	4	45.38	312.08	0.500
	5	11	45.46	338.98	0.364
	6	9	43.43	301.57	0.333
Total		47	43.02	320.28	0.234

(a) BW: birth weight in kg.; PO: pelvic opening in cm^2; CD: frequency of difficult calvings.

The model for pelvic opening was

$$Z_{ijkl} = D'_i + T'_j + S'_k + e'_{ijkl} \qquad (56\,a)$$

where D'_i is the effect of the i^{th} department of origin of the heifer ($i = 1,2$), T'_j is the effect of the j^{th} season of calving ($j = 1,2$), S'_k is the effect of the k^{th} sire of heifer ($k = 1, \ldots, 6$) and e'_{ijkl} is a residual. The vectors $\boldsymbol{\beta}_2$ and $\mathbf{u}_2$ were defined as

$$\boldsymbol{\beta}'_2 = [D'_1 + T'_2,\ D'_2 + T'_2,\ T'_1 - T'_2] \qquad (56\,b)$$

$$\mathbf{u}'_2 = [S'_1,\ S'_2,\ S'_3,\ S'_4,\ S'_5,\ S'_6]. \qquad (56\,c)$$

The data in Table 1 can be regarded as a 47×2 contingency table, with rows corresponding to each record, and columns being « DIFFICULT » and « EASY » calvings. Hence, $n_{i.} = 1$ for $i = 1, \ldots, 47$, and $\mathbf{Y}' = [Y_1, \ldots, Y_{47}]$, with Y_i being a scalar variable with realized value 1 if a difficult calving occurs, or 0 otherwise. The probability of *difficult* calving for the i^{th} row was assumed a normal integral with argument modeled as

$$\mu_{i(jklm)} = D''_j + T''_k + L''_l + S''_m + b_1(Y_{ijklm} - 43.02) + b_2(Z_{ijklm} - 320.28) \qquad (57\,a)$$

where D''_j is the effect of the j^{th} department of origin ($j = 1,2$), T''_k is the effect of the k^{th} season of calving ($k = 1,2$), L''_l is the effect of the l^{th} sex ($l = 1$: male, $2 =$ female), and S''_m is the effect of the m^{th} sire of the heifer; b_1 and b_2 are partial « regression » coefficients of the underlying variate on birth weight of the calf and pelvic opening of the heifer, respectively. These coefficients were assumed known with $b_1 = .1643$ and $b_2 = -.0184$; the logic for the choice of these values is presented in the following section. Note that as $\mu_{i(jklm)}$ increases, so does the probability of difficult calving; also, $\mu_{i(jklm)}$

increases with increased birth weight and decreases with increased pelvic opening. The vector $\boldsymbol{\tau}$ and $\boldsymbol{\nu}$ were then

$$\boldsymbol{\tau} = [D''_1 + T''_2 + L''_2,\ D''_2 + T''_2 + L''_2,\ T''_1 - T''_2,\ L''_1 - L''_2] \qquad (57\,b)$$

$$\boldsymbol{\nu} = [S''_1,\ S''_2,\ S''_3,\ S''_4,\ S''_5,\ S''_6). \qquad (57\,c)$$

B. *Conditional covariance*

Given $\boldsymbol{\theta}$, the variance-covariance matrix of birth weight and pelvic opening is

$$\text{Var}\begin{bmatrix} \mathbf{y}_1|\boldsymbol{\theta} \\ \mathbf{y}_2|\boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} \sigma^2_{e_1} & \sigma_{e_{12}} \\ \sigma_{e_{12}} & \sigma^2_{e_2} \end{bmatrix} \otimes \mathbf{I}_{47\times 47}$$

where $\otimes$ is the Kronecker product. The values used for the residual covariance matrix were (MENISSIER & SAPA, personal communication): $\sigma^2_{e_1} = 25$, $\sigma^2_{e_2} = 1089$ and $\sigma_{e_{12}} = 41.25$.

The coefficients b_1 and b_2 were calculated as in (16) and (17) from $\rho_{12} = .25$, $\rho_{13} = .50$ and $\rho_{23} = -.30$; the residual variance in the underlying scale, which was set equal to 1, corresponds to (15). These values yielded $b_1 = .1643$ and $b_2 = -.0184$.

C. *Prior distribution*

The parameter vector for this problem was

$$\boldsymbol{\theta}' = [\boldsymbol{\beta}'_1 \boldsymbol{\beta}'_2 \boldsymbol{\tau}' \mathbf{u}'_1 \mathbf{u}'_2 \boldsymbol{\nu}']. \qquad (58)$$

Prior knowledge about $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\tau}$ was assumed to be vague. The covariance matrix of $\mathbf{u}_1$, $\mathbf{u}_2$, and $\boldsymbol{\nu}$ was

$$\text{Var}\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \boldsymbol{\nu} \end{bmatrix} = \mathbf{G}_c \otimes \mathbf{I}_{6\times 6} \qquad (59)$$

where $\mathbf{G}_c$ is a 3×3 matrix calculated as in (31). The unconditional prior covariance matrix was taken as

$$\mathbf{G} = \{\rho_{G_{ij}} \sigma_{u_i} \sigma_{u_j}\} \quad \begin{matrix} i = 1, \ldots, 3 \\ j = 1, \ldots, 3 \end{matrix} \qquad (60)$$

where $\rho_{G_{ij}}$ is the genetic correlation between traits i and j in the underlying scale. The genetic correlations used were (MENISSIER & SAPA, personal communication) : $\rho_{G_{13}} = .70$ and $\rho_{G_{23}} = -.50$. The standard deviations were calculated as

$$\sigma_{u_i} = \sigma_{e_i}/\sqrt{\lambda_i}, \quad i = 1, \ldots, 3 \qquad (61)$$

with $\lambda_i = (4 - h^2_i)/h^2_i$, and $h^2_1 = .15$, $h^2_2 = .40$ and $h^2_3 = .30$. Further

$$\sigma_{e_3} = 1/\sqrt{1 - \rho^2_{3.12}} \qquad (62)$$

with $\rho^2_{3.12} = .4427$. We obtained

$$\mathbf{G} = \begin{bmatrix} .9740 & 3.7997 & .2635 \\ 3.7997 & 121.0000 & -2.0978 \\ .2635 & -2.0978 & .1455 \end{bmatrix}; \quad \mathbf{G}_c = \begin{bmatrix} .9740 & 3.7997 & .1734 \\ 3.7997 & 121.0000 & -.4956 \\ .1734 & -.4956 & .0260 \end{bmatrix}$$

Computations were also carried under the hypothesis of no «risk» relationship, i.e., $b_1 = b_2 = 0$. In this case, a different prior covariance matrix was used

$$\mathbf{G}_0 = \begin{bmatrix} .9740 & 3.7997 & .1967 \\ 3.7997 & 121.0000 & -1.5661 \\ .1967 & -1.5661 & .0811 \end{bmatrix}$$

obtained from **G** by appropriate rescaling of elements. For example, and taking into account that $1/\sqrt{1-\rho^2_{3.12}} = 1.3395$

$$.0811 = .1455/(1.3395)^2.$$

Note that $h^2_3 = 4 \times .0811/(1 + .0811) = .30$, $\rho_{G_{13}} = .70$ and $\rho_{G_{23}} = -.50$, as it should be. In this instance, the μ_i's are expressed in standard deviation units of the underlying variate for calving difficulty «unadjusted» for residual variation in birth weight and pelvic opening. In order to compare estimates obtained under $b_1 \neq 0$ and $b_2 \neq 0$ with those calculated with $b_1 = b_2 = 0$, the latter were multiplied by 1.3395 to express them in the same scale.

D. *Logistic approximation*

In each of the two cases ($b_1 \neq 0$ and $b_2 \neq 0$, and $b_1 = b_2 = 0$) computations were also conducted using the logistic approximation in (11 *b*). Since the residual variance in the logistic scale is $\pi^2/3$, the prior covariance matrices $\mathbf{G}_c$ and $\mathbf{G}_0$ discussed in the previous section were rescaled as

$$\mathbf{G}^*_c = \mathbf{L}\mathbf{G}_c\mathbf{L}'; \quad \mathbf{G}^*_0 = \mathbf{L}\mathbf{G}_0\mathbf{L}'$$

where **L** is a 3×3 diagonal matrix with elements 1, 1 and $\pi/\sqrt{3}$. Solutions to (45) and (46) obtained with the logistic approximation were then divided by $\pi/\sqrt{3}$ to make them comparable to those obtained with the normal scale.

E. *Iteration*

Starting values for $\boldsymbol{\tau}$ and $\boldsymbol{\nu}$ are needed to iterate with (45) or (46). Two different sets of starting values were used. The first was the $\boldsymbol{\tau}$ and $\boldsymbol{\nu}$ roots of (45) with $\mathbf{W}^{[i-1]} = \mathbf{I}$, $\mathbf{v}^{[i-1]} = \mathbf{t}$ being a vector of (0,1) variables (1: difficult calving; 0: otherwise) and $\boldsymbol{\nu}^{[i-1]} = \mathbf{0}$. These roots yielded $\boldsymbol{\tau}^{[0]}$ and $\boldsymbol{\nu}^{[0]}$ which were used to compute $\mu^{(0)}_{i(jklm)}$ in (57 *a*); in turn, these values permitted calculation of $\mathbf{W}^{(0)}$ in both the normal and logistic cases. The second starting set was the solution to (45) with $\mathbf{W}^{[i-1]} = \mathbf{I}$, $\mathbf{v}^{[i-1]} = \mathbf{t}^*$ being a vector of empirical logits $\left(\ln\left[\frac{1+.5}{0+.5}\right] = 1.099\right.$ if a difficult calving occurred and -1.099 otherwise) and $\boldsymbol{\nu}^{[i-1]} = \mathbf{0}$.

Iteration stopped when $\sqrt{\Delta'\Delta/29} < 10^{-10}$, where $\boldsymbol{\Delta} = \hat{\boldsymbol{\theta}}^{[i]} - \hat{\boldsymbol{\theta}}^{[i-1]}$. In each of the four cases resulting from the combination of normal or logistic functions with hypotheses about residual correlation ($b_1 \neq 0$ and $b_2 \neq 0$ vs, $b_1 = b_2 = 0$), convergence to the same solution occurred irrespective of the starting set used. Six rounds of iteration were required for the starting set using $\mathbf{v}^{[i-1]} = \mathbf{t}^*$; seven rounds were required when $\mathbf{v}^{[i-1]} = \mathbf{t}$ was used. From a practical point of view, however, iteration could have stopped at the third round. Results of iteration using a normal integral, $b_1 \neq 0$ and $b_2 \neq 0$, and $\mathbf{v}^{[i-1]} = \mathbf{t}$ as a trial vector are shown in Table 3.

TABLE 3

Solutions by round of iteration in the case of a normal function, $b_1 \neq 0$ and $b_2 \neq 0$, and a trial vector of (0, 1) variables.

Solution à différentes itérations dans le cas d'une fonction normale avec $b_1 \neq 0$, $b_2 \neq 0$ et un vecteur initial de variables (0, 1).

Component[a] of $\boldsymbol{\theta}$	Iteration[b]		
	0	1	7
1. $D_1 + T_2 + L_2$	41.664	41.671	41.688
2. $D_2 + T_2 + L_2$	42.307	42.282	42.315
3. $T_1 - T_2$	1.221	−1.208	−1.194
4. $L_1 - L_2$	3.028	3.017	2.978
5. S_1	−0.336	−0.305	−0.426
6. S_2	−0.210	−0.243	−0.270
7. S_3	−0.633	−0.582	−0.664
8. S_4	0.372	0.352	0.491
9. S_5	0.730	0.821	0.885
10. S_6	0.077	−0.043	−0.017
11. $D_1' + T_2'$	313.138	313.134	313.143
12. $D_2' + T_2'$	312.590	312.644	312.610
13. $T_1' - T_2'$	14.899	14.876	14.848
14. S_1'	4.401	4.349	4.548
15. S_2'	0.168	0.228	0.284
16. S_3'	−3.800	−3.907	−3.738
17. S_4'	−1.855	−1.812	−2.106
18. S_5'	11.113	10.968	10.876
19. S_6'	−10.027	−9.826	−9.864
20. $D_1'' + T_2'' + L_2''$	0.135	−1.193	−1.772
21. $D_2'' + T_2'' + L_2''$	0.107	−1.390	−2.134
22. $T_1'' - T_2''$	−0.041	0.405	0.432
23. $L_1'' - L_2''$	0.244	0.420	1.022
24. S_1''	−0.116	−0.113	−0.126
25. S_2''	−0.050	−0.053	−0.056
26. S_3''	−0.084	−0.078	−0.088
27. S_4''	0.091	0.088	0.106
28. S_5''	0.030	0.039	0.045
29. S_6''	0.129	0.116	0.119

(a) Components 1-10: birth weight, 11-19: pelvic opening, 20-29: underlying variate for calving difficulty.

(b) Convergence attained at the seventh round of iteration.

F. *Model fit, estimates and their posterior precision*

The models were evaluated for fit by referring the statistic in (51) to a chi-square distribution with $47 - 4 = 43$ degrees of freedom. None of the chi-square values

$$\chi^2(b_1 \neq 0,\ b_2 \neq 0,\ \text{normal}) = 26.19; \quad \chi^2(b_1 = b_2 = 0,\ \text{normal}) = 37.56$$

$$\chi^2(b_1 \neq 0,\ b_2 \neq 0,\ \text{logistic}) = 27.12; \quad \chi^2(b_1 = b_2 = 0,\ \text{logistic}) = 37.35$$

could be considered significant so there was no evidence to reject the model. However,

given the sparsity of the contingency table analyzed in this example, the approximation of (51) to a chi-square statistic may be poor.

Differences between final round estimates of $\boldsymbol{\theta}$ obtained with the normal ($\widehat{\boldsymbol{\theta}}_N$) and the logistic ($\widehat{\boldsymbol{\theta}}_L$) functions were small so the latter will not be presented here. In fact,

$$\left[(\widehat{\boldsymbol{\theta}}_L - \widehat{\boldsymbol{\theta}}_N)'(\widehat{\boldsymbol{\theta}}_L - \widehat{\boldsymbol{\theta}}_N)/29\right]^{\frac{1}{2}} = 2.94 \times 10^{-2}.$$

Estimates of components of $\boldsymbol{\theta}$ obtained using the normal distribution, and their estimated posterior precision (square root of estimated posterior variance) are shown in Table 4. The contrast $L_1'' - L_2''$ was estimated at 1.022 and 1.315 for the cases ($b_1 \neq 0$, $b_2 \neq 0$) and ($b_1 = b_2 = 0$), respectively. These indicate that if a male calf is born, the probability of a « difficult » calving would be larger than if a female calf is born,

TABLE 4

Estimates of components of $\boldsymbol{\theta}$ and their posterior precision in the case of a normal function.

Estimations des composantes de $\boldsymbol{\theta}$ et de leur précision a posteriori dans le cas d'une fonction normale.

Component[(a)] of $\boldsymbol{\theta}$	« Risk » relationship	
	$b_1 \neq 0,\ b_2 \neq 0$	$b_1 = 0,\ b_2 = 0$
1. $D_1 + T_2 + L_2$	41.688 ±1.425	41.697 ±1.424
2. $D_2 + T_2 + L_2$	42.315 ±1.669	42.378 ±1.669
3. $T_1 - T_2$	−1.194 ±1.502	−1.206 ±1.502
4. $L_1 - L_2$	2.978 ±1.457	2.937 ±1.455
5. S_1	−0.426 ±0.862	−0.571 ±0.847
6. S_2	−0.270 ±0.867	−0.240 ±0.854
7. S_3	−0.664 ±0.894	−0.805 ±0.884
8. S_4	0.491 ±0.899	0.544 ±0.891
9. S_5	0.885 ±0.831	0.878 ±0.815
10. S_6	−0.017 ±0.852	0.194 ±0.834
11. $D_1' + T_2'$	313.143 ±8.468	313.287 ±8.465
12. $D_2' + T_2'$	312.610 ±10.710	312.362 ±10.707
13. $T_1' - T_2'$	14.848 ±9.970	14.756 ±9.968
14. S_1'	4.548 ±8.361	5.605 ±8.306
15. S_2'	0.284 ±8.763	0.314 ±8.672
16. S_3'	−3.738 ±9.125	−2.422 ±9.059
17. S_4'	−2.106 ±9.387	−3.268 ±9.297
18. S_5'	10.876 ±8.235	10.313 ±8.146
19. S_6'	−9.864 ±8.525	−10.543 ±8.433
20. $D_1'' + T_2'' + L_2''$	−1.772 ±0.563	−1.666 ±0.634
21. $D_2'' + T_2'' + L_2''$	−2.134 ±0.692	−1.957 ±0.777
22. $T_1'' - T_2''$	0.432 ±0.522	−0.150 ±0.599
23. $L_1'' - L_2''$	1.022 ±0.588	1.315 ±0.643
24. S_1''	−0.126 ±0.129	−0.361 ±0.310
25. S_2''	−0.056 ±0.133	−0.100 ±0.311
26. S_3''	−0.088 ±0.138	−0.217 ±0.326
27. S_4''	0.106 ±0.141	0.285 ±0.328
28. S_5''	0.045 ±0.124	0.016 ±0.294
29. S_6''	0.119 ±0.128	0.376 ±0.302

(a) Components 1-10: birth weight; 11-19: pelvic opening; 20-29: underlying variate for calving difficulty.

irrespective of whether the effects of birth weight and pelvic opening are removed. This is consistent with the findings of BELIC & MENISSIER (1968). However, the difference in the underlying scale between male and female calves was smaller when birth weight was included as a « risk » variable. If this result were true, it would suggest that part of the difference between sexes in liability for calving difficult is not associated with differences in birth weight. The effect of including « risk » variables in the model was clear in relation to differences between seasons. Season 1 was more favourable in the ($b_1 = 0$, $b_2 = 0$) model perhaps because of calves with lighter birth weight and dams with larger pelvic opening; when these differences were taken into account ($b_1 \neq 0$, $b_2 \neq 0$), season 2 turned out to be more favourable.

G. *Sire evaluation*

As pointed out before, $\boldsymbol{\nu} = \mathbf{u}_3 - b_1\mathbf{u}_1 - b_2\mathbf{u}_2$, so sire solutions presented in Table 4 for the two different models are not comparable. Sires can be ranked for calving difficulty in the full model by using the statistic

$$\hat{\mathbf{u}}_3 = \hat{\boldsymbol{\nu}} + b_1\hat{\mathbf{u}}_1 + b_2\hat{\mathbf{u}}_2 \tag{63}$$

where $\hat{\boldsymbol{\nu}}$, $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ are the sire components of $\hat{\boldsymbol{\theta}}$ associated with the underlying variate, birth weight and pelvic opening, respectively. From a practical point of view, one may be interested in ranking sires in terms of probability of difficult calving rather than in a hypothetical underlying scale. For example, breeders may wish to know the probability that a heifer sired by the m^{th} bull, born in region 1, calving a male calf in season 1 a will experience a difficult calving. An estimate of this probability can be calculated as

$$\hat{\Pi}_{111m} = \Phi\left[\hat{D}''_1 + \hat{T}''_1 + \hat{L}'_1 + S''_m + b_1(\hat{D}_1 + \hat{T}_1 + \hat{L}_1 - 43.02) + b_2(\hat{D}'_1 + \hat{T}'_1 - 320.28)\right]. \tag{64}$$

Using (64) for sires 1 to 6 yields

Sire 1: .253 Sire 4: .436

Sire 2: .312 Sire 5: .347

Sire 3: .304 Sire 6: .463.

In more general situations, e.g., artificial insemination, the probability of difficult calving associated with using the m^{th} sire in a given distribution of regions, calving seasons and sexes of calf may be of interest. This probability could be estimated as

$$\hat{\Pi}_{\cdots m} = \Sigma_{jkl}\delta_{jkl}\hat{\Pi}_{jklm} \tag{65}$$

with $\hat{\Pi}_{jklm}$ as in (64) and δ_{jkl} being an arbitrary weight such that $\Sigma_{jkl}\delta_{jkl} = 1$. For the example considered in this paper, we took $\delta = 1/8$ because there were 8 region × season × sex subclasses, and ranked sires using (63) and (65). Results are shown in Table 5 for the normal and logistic distributions. As already indicated, differences between the normal and logistic models were negligible, and the estimated probability of difficult calving ranged between .116 and .239. Note that evaluations based on raw frequencies (Table 2) gave the probability rankings :

Sire 1 = Sire 3 < Sire 2 < Sire 6 < Sire 5 < Sire 4.

However, the ranking in Table 5 was

Sire 1 < Sire 3 < Sire 2 < Sire 5 < Sire 4 < Sire 6.

This indicates that evaluation based on raw frequencies can be seriously misleading. However, the progeny group sizes were small (Table 2) and none of the evaluations calculated with (63) could be considered different from zero (Table 5).

TABLE 5

Sire evaluation and estimated probability of difficulty calving[(a)].

Évaluation des pères et estimation de la probabilité d'apparition d'un vêlage difficile.

Sire	Evaluation ± Posterior precision		Probability of Difficult calving	
	Normal	Logistic	Normal	Logistic
1	−0.280±0.321	−0.279±0.321	0.117	0.116
2	−0.106±0.326	−0.103±0.326	0.147	0.148
3	−0.128±0.337	−0.127±0.336	0.143	0.144
4	0.225±0.341	0.232±0.340	0.217	0.226
5	−0.009±0.311	−0.007±0.310	0.166	0.169
6	0.298±0.319	0.284±0.318	0.235	0.239

(a) See equations (63) and (65).

VII. Conclusions

This paper presents a solution to the problem of estimating the genetic merit of candidates for selection when both quantal and continuous information is available in a set of individuals. The proposed method was adapted to the situation where the probability of « response » is a function of continuous « risk » variables. Also, consideration is given to the assumption that candidates for selection are sampled from a distribution with second moments known, a priori. The method can be extended to multiple ordered or unordered categories of response along the lines presented by GIANOLA & FOULLEY (1983).

The method is non-linear and approximates the best predictor in a squared error sense. Theoretical objections arising in analysis of categorical data with linear models (e.g., GIANOLA, 1982) are eliminated. For example, when calving difficulty is measured as an « all or none » trait, sire × sex of calf interactions are usually found to be « significant ». This may be associated with a scaling problem. Suppose we wish to compare two sires and that the values in the underlying scales are μ_{1M}, μ_{1F}, μ_{2M} and μ_{2F}; the subscripts indicate the sire and the sex of the calf. Further, suppose that there is no interaction between sex and sire in the underlying scale, i.e.,

$$\mu_{1M} - \mu_{2M} = \mu_{1F} - \mu_{2F}.$$

However, $\Phi(\mu_{1M}) - \Phi(\mu_{2M})$ may be different from $\Phi(\mu_{1F}) - \Phi(\mu_{2F})$ because $\Phi(x)$ does not vary linearly with x.

The method of estimation is based on Bayes theorem, but is not completely Bayesian in the sense that the variance-covariance structure is regarded as representing a set of « nuisance » parameters. In principle, prior knowledge (or lack of) about variances and covariances could be represented via a prior distribution (LINDLEY & SMITH, 1972) and modal estimates obtained from the posterior density. HARVILLE (1977) has indicated that estimators of variances obtained from the joint posterior mode can be degenerate if uninformative priors are used. This author qualified the modes of the marginal posterior density of the variance components as « seemingly superior » estimators.

Important numerical problems arise when the procedure is applied to the estimation of vectors with thousands of elements, the usual situation in applied animal breeding. Nevertheless, the order of the computations is comparable to that arising in multi-dimensional BLUP multiplied by the number of « main » iterates needed to achieve convergence. When the « risk » variables are considered in the model, the method requires that every experimental unit with a categorical response includes information on the quantitative variates.

Acknowledgements

Daniel GIANOLA wishes to acknowledge I.N.R.A., France, for support during his stay at Jouy-en-Josas, and the Holstein Association, Brattleboro, Vermont, U.S.A., for supporting his work in categorical data. Dr. Stephen P. SMITH, Cornell University, U.S.A., is thanked for useful comments.

Received February 2, 1983.
Accepted April 29, 1983.

References

BECKMAN F.S., 1960. *The solution of linear equations by the conjugate-gradient method. In :* Ralston A., Wilf H.S., *Mathematical methods for digital computers,* Wiley, New York.

BELIC M., MENISSIER F., 1968. Étude de quelques facteurs influençant les difficultés du vêlage en croissement industriel. *Ann. Zootech.,* **17**, 107-142.

CORNFIELD J., 1969. The Bayesian outlook and its application. *Biometrics,* **25**, 617-657.

COX D.R., 1970. *The Analysis of Binary Data.* Chapman and Hall, London.

DAHLQUIST G., BJÖRCK A., 1974. *Numerical Methods.* Prentice Hall, Englewood Cliffs.

DEMPFLE L., 1977. Relation entre BLUP (Best Linear Unbiased Prediction) et estimateurs Bayésiens. *Ann. Genet. Sel. Anim.,* **9**, 27-32.

DEMPSTER E.R., LERNER I.M., 1950. Heritability of threshold characters. *Genetics,* **35**, 212-235.

FINNEY D.J., 1952. *Probit analysis : a statistical treatment of the sigmoïd response curve.* 2nd ed., University Press, Cambridge.

GIANOLA D., 1980. A method of sire evaluation for dichotomies. *J. Anim. Sci.,* **51**, 1266-1271.

GIANOLA D., 1982. Theory and analysis of threshold characters. *J. Anim. Sci.,* **56**, 1079-1096.

GIANOLA D., FOULLEY J.L., 1983. Sire evaluation for ordered categorical data with a threshold model. *Génét. Sél. Evol.,* **15**, 201-223.

HARVILLE D.A., 1977. Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.,* **72**, 320-338.

HENDERSON C.R., 1973. Sire evaluation and genetic trends. *Proceedings of the Animal Breeding and Genetics Symposium in honor of Dr Jay L. Lush, Blacksburg, Virginia, July 29, 1972,* 10-41, ASAS-ADSA, Champaign, Illinois.

HENDERSON C.R., 1980. A simple method for unbiased estimation of variance components in the mixed model. *J. Anim. Sci.,* **51** (Suppl. 1), 119.

HENDERSON C.R., KEMPTHORNE O., SEARLE S.R., VON KROSIGK C.M., 1959. The estimation of environmental and genetic trends from records subject to culling. *Biometrics,* **15**, 192-218.

HILL B.M., 1965. Inference about variance components in the one-way model. *J. Am. Stat. Assoc.,* **60**, 806-825.

LEONARD T., 1972. Bayesian methods for binomial data. *Biometrika,* **59**, 581-589.

LINDLEY D.V., SMITH A.F.M., 1972. Bayes estimates for the Linear model. *J.R. Stat. Soc. B.,* **24**, 1-41.

MOOD A.M., GRAYBILL F.A., 1963. *Introduction to the Theory of Statistics.* McGraw-Hill Book Co., New York.

RÖNNINGEN K., 1971. Some properties of the selection index derived by « Henderson's Mixed Model Method ». *Z. Tierz. Züchtbiol.*, **88**, 186-193.

SCHAEFFER L.R., WILTON J.W., THOMPSON R., 1978. Simultaneous estimation of variance and covariance components from multitrait mixed model equations. *Biometrics*, **34**, 199-208.

THOMPSON R., 1979. Sire evaluation. *Biometrics*, **35**, 339-353.

TIAO G.C., TAN W.Y., 1965. Bayesian analysis of random effects models in the analysis of variance. I. Posterior distribution of variance components. *Biometrika*, **52**, 37-53.

TIAO G.C., BOX G.E.P., 1967. Bayesian analysis of a three-component hierarchical design model. *Biometrika*, **54**, 109-125.

VAN NORTON R., 1960. The solution of linear equations by the Gauss-Seidel Method. *In :* Ralston A. Wilf H.S., *Mathematical methods for digital computers*, Wiley, New York.

30

Composite link functions in generalised linear models

by

R. Thompson and R.J. Baker

Composite Link Functions in Generalized Linear Models

BY

R. THOMPSON and R. J. BAKER

Reprinted from

THE JOURNAL OF THE ROYAL STATISTICAL SOCIETY

SERIES C (APPLIED STATISTICS)

Volume 30, No. 2, 1981

(p. 125)

PRINTED FOR PRIVATE CIRCULATION

1981

Appl. Statist. (1981),
30, *No.* 2, *pp.* 125–131

Composite Link Functions in Generalized Linear Models

By R. THOMPSON and R. J. BAKER

A.R.C. Unit of Statistics, University of Edinburgh Scotland, UK

Rothamsted Experimental Station, Herts, UK

[Received January 1979. Final revision November 1980]

SUMMARY

In generalized linear models each observation is linked with a predicted value based on a linear function of some systematic effects. We sometimes require to link each observation with a linear function of more than one predicted value. We embed such models into the generalized linear model framework using composite link functions. The computer program GLIM-3 can be used to fit these models. Illustrative examples are given including a mixed-up contingency table and grouped normal data.

Keywords: GLIM; MIXED-UP TABLES; EM ALGORITHMS; GROUPED NORMAL DATA; RELATIVE RISK

1. INTRODUCTION

GENERALIZED linear models (Nelder and Wedderburn, 1972) have proved a useful unifying concept for identifying data structure and have allowed a flexible approach to model fitting. The generalized linear model is specified by

(i) Independent observations $y_1, ..., y_n$ distributed according to an exponential family distribution.

(ii) A set of explanatory variables $x_{i1}, ..., x_{ip}$, available on each observation ($i = 1, ..., n$), describing the systematic linear component through $\eta_i = \Sigma x_{ij} \beta_j$, summing over j from 1 to p.

(iii) The link function $g(\mu_i) = \eta_i$ relating the mean of an observation to the linear predictor η_i.

For example, consider a two-way contingency table with independence. The data, in the form of counts, can be thought of as arising from a Poisson distribution (for example, Birch, 1963). With independence, the mean of each observation (μ_i) can be thought of as a row effect times a column effect. This means that a logarithmic link function $\log(\mu_i) = \eta_i$ allows a linear structure to be defined linking the mean of an observation to explanatory variables representing rows and columns of the table.

Although generalized linear models are very flexible, some types of data do not fit directly into this framework. As an example consider the O, A, B system for human blood groups (Kempthorne, 1969, p. 172). Suppose that the respective gene frequencies are r, p and q and that there is random mating with respect to this locus. Then the frequencies in the four observed groups O, A, B and AB are r^2, $p^2 + 2pr$, $q^2 + 2qr$ and $2pq$ respectively. One can think of the frequencies as arising from a "mixed-up" contingency table in which the AA, AO and OA cells are indistinguishable and similarly the BB, BO and OB cells, and the AB and BA cells are mixed-up.

If instead of linking each observation with one linear predictor we link a composite function of more than one linear predictor with each observation we can embed mixed-up contingency tables into the generalized linear model framework. In Section 2 we develop the necessary

Frequency table for O, A, B system

		A	O	B
		p	r	q
A	p	p^2	pr	pq
O	r	rp	r^2	rq
B	q	pq	qr	q^2

theory for such composite link functions. In Section 3 we show, using numerical examples, that this extension is useful not only for mixed-up contingency tables but in other contexts including grouped normal data and polychotomous responses. A further extension to the composite link function when the data can be usefully represented by the combination of two linear models is discussed briefly in Section 4. More general forms of composite link functions are considered in Section 5.

In Section 6 we discuss the efficiency of the procedure. In Section 7 some comments about computation are made including a GLIM-3 program for the contingency table example.

2. Theory

The relationship between the mean of an observation and its linear predictor in generalized linear models can be written as $g(\mu_i) = \eta_i$ or alternatively $\mu_i = h(\eta_i)$. Sometimes we wish to associate more than one η_k (or correspondingly more than one set of systematic effects x_{kj} $(j = 1, ..., p)$) with each observation y_i. We suppose that $\gamma_k = h(\eta_k)$ $(k = 1, ..., m)$, where m is the number of sets of systematic effects, and that $\mu_i = c_i(\boldsymbol{\gamma})$ where the c_i are known functions and $\boldsymbol{\gamma}$ is a $p \times 1$ vector with elements γ_k. Since each c_i can link μ_i, an expected value, with several η_k, functions of several linear predictors, we call c_i a *composite link function*. For ease of presentation we will now consider the special case when μ_i is a known linear function of $\boldsymbol{\gamma}$ and relegate the more general case to Section 5. Then μ_i can be written as $\mu_i = \Sigma c_{ik} \gamma_k$, summing over k from 1 to m. The model can be written as $E(\mathbf{y}) = \boldsymbol{\mu} = C\boldsymbol{\gamma}$, $\boldsymbol{\gamma} = h(\boldsymbol{\eta})$, $\boldsymbol{\eta} = X\boldsymbol{\beta}$, where $\boldsymbol{\mu}$, $\mathbf{y}$ are $n \times 1$ vectors, $\boldsymbol{\beta}$ is a $p \times 1$ vector and X is a $m \times p$ matrix with respective elements μ_i, y_i β_i and x_{ij}. In this, C is a $n \times m$ matrix, the matrix analogue of the composite link functions.

Nelder and Wedderburn (1972) show that for their generalized linear models the solution of the likelihood equations for β_j $(j = 1, ..., p)$ is equivalent to an iterative weighted least-squares procedure. In each iteration a modified dependent variable given by $z_i = \eta_i + (y_i - \mu_i).(d\eta_i/d\mu_i)$ is used and the weight given to z_i is $w_i = (d\mu_i/d\eta_i)^2/\tau_i^2$ where $\tau^2 = d\mu/d\theta$ and θ is a parameter of the density function of y. For instance $\tau^2 = 1$ and μ for the Normal and Poisson distributions.

For our model with a composite matrix, C, a simple modification of Nelder and Wedderburn's iterative procedure gives maximum likelihood estimates of $\boldsymbol{\beta}$. In each iteration working dependent and independent variables $\mathbf{z} = CH\boldsymbol{\eta} + (\mathbf{y} - \boldsymbol{\mu})$ and $X^* = CHX$ are used in least-squares equations with weight function $1/\tau_i^2$ for z_i. Here, H is a $m \times m$ diagonal matrix with diagonal elements equal to $d\gamma_k/d\eta_k$. Obviously when $C = I$ and $\mu_k = \gamma_k$ this procedure is equivalent to that of Nelder and Wedderburn.

Since the inverse of a composite link function does not always exist we cannot, as Nelder and Wedderburn suggest for simple link functions, take the observations as initial estimates of the expected value (μ) and find estimates in the linear predictors $(\boldsymbol{\eta})$ through the inverse of the *Ch.* However, it is usually a simple matter to find satisfactory first estimates of $\boldsymbol{\eta}$ by using the values found from fitting a similar model with a simple link function.

3. Examples of Fitting Composite Link Function Models

We now illustrate the fitting of models incorporating composite link functions by considering three examples.

3.1. *Mixed-up Contingency Tables*

We take as an example the O, A, B system mentioned in the Introduction, with data given by Taylor and Prior (1938) and used by Kempthorne (1969). They observed 202, 179, 35 and 6 people in the 4 groups O, A, B and AB. The model fitted to the genotypes is additive on the logarithmic scale so that $h(\eta) = \exp(\eta)$. If we write $\mathbf{exp}(\boldsymbol{\eta})$ and $\mathbf{ln}(\boldsymbol{\eta})$ as the vectors with kth elements $\mathbf{exp}(\boldsymbol{\eta}_k)$ and $\mathbf{ln}(\boldsymbol{\eta}_k)$, the model fitted is of the form $\mu = C\mathbf{h}(X\boldsymbol{\beta}) = C\,\mathbf{exp}(X\boldsymbol{\beta})$ where X and $\boldsymbol{\beta}$ are respectively 9×4 and 4×1 matrices and C, the matrix analogue of the composite link function, is a 4×9 matrix. Write $\boldsymbol{\beta} = \mathbf{ln}(\boldsymbol{\alpha})$ where $\boldsymbol{\alpha}$ is a 4×1 vector given by $\boldsymbol{\alpha}^t = (N, p, r, q)^t$. Here, N is the total number of observations, and p, r, q are the frequencies of A, O and B.

The matrices X, C and $\mathbf{y}$ are given by

$$X^t = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 2 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 2 \end{bmatrix},$$

$$C = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

and $y^t = (179\ 6\ 202\ 35)$.

In this example, following Birch (1963), we think of the observations as coming from independent Poisson distributions so that $\tau_i^2 = \mu_i$; also since $h = \exp$, $d\gamma_k/d\gamma_k = \exp(\eta_k)$. There are dependencies in the columns of the design matrix X and just as in ordinary linear models we need to introduce constraints. We use the constraint $\beta_1 = 0$.

Using initial estimates of $\beta_j = \frac{1}{2}\ \ln(422/9)$ $(j = 2, 3, 4)$ convergence was attained in 5 iterations. The fitted values for the O, A, B and AB were 205·9, 175·0, 30·6 and 10·6 respectively. Hence estimates of p, r and q are 0·252, 0·698 and 0·050. The deviance is 3·17 with one degree of freedom. The GLIM-3 coding for this example is given in Section 7.

3.2. *Grouped Normal Data*

The exact analysis of grouped normal data has been presented by (among others) Sampford (1952). We apply the method to data taken from Snedecor and Cochran (1967, p. 105) which refer to gains in weight of two samples of female rats under two diets. For the purposes of this example we round the data to the nearest 10 g so that they now appear as

Diet 1

Weight gain	0–85	–95	–105	–115	–125	–135	–145	–155	155–
Frequency	1	0	2	2	3	2	0	1	1

Diet 2

Weight gain	0–75	–85	–95	–105	–115	–125	125–
Frequency	1	0	2	1	1	1	1

The original data are assumed to be normally distributed with means μ_1 and μ_2 for the two diet levels and a common variance σ^2. Hence for each level the observed frequencies (y_i) are

treated as Poisson variables. For the first diet $E(y_{1,i}) = \mu_{1,i}$, $\mu_{1,1} = N_1 \Phi(\eta_{1,1})$, $\mu_{1,i} = N_1[\Phi(\eta_{1,i}) - \Phi(\eta_{1,i-1})]$ $(i > 1)$ where Φ is the cumulative normal integral and $\eta_{1,i} = (-\mu_1 + u_{1,i})/\sigma$ where $u_{1,i}$ is the upper bound for the ith class, and similarly for diet 2. In this case the composite link function links $\mu_{1,i}$ with $\eta_{1,i}$ and $\eta_{1,i-1}$ $(i > 1)$. Strictly N_1 and N_2 are parameters that require to be estimated but with data of this form obvious estimates are the sample sizes, i.e. $N_1 = 12, N_2 = 7$. In the next section we show how estimation of N_1 and N_2 can be put into the linear model framework with a consequent reduction by two in the residual degrees of freedom. In GLIM we specify

$$C = \begin{bmatrix} N_1 E_1 & O_{9,7} \\ O_{7,9} & N_2 E_2 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} -\mu_1/\sigma \\ 1/\sigma \\ -\alpha/\sigma \end{bmatrix},$$

$$X^t = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & & & & & & & \\ 85 & 95 & 105 & 115 & 125 & 135 & 145 & 155 & 1000 & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & & & & & & \\ & & & & & & & & & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ & & & & & & & & & 75 & 85 & 95 & 105 & 115 & 125 & 1000 \\ & & & & & & & & & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

The E_t are square matrices with ones on the diagonal, minus ones on the subdiagonal and zeros elsewhere, E_1 of order 9×9 and E_2 of order 7×7, $O_{u \times v}$ is a matrix of zeros of order $u \times v$ and $\alpha = \mu_2 - \mu_1$. The function h is the cumulative normal integral. In this example note that although the number of observations (n) equals the number of linear predictors (m), μ_i is not a function of η_i alone.

Initial estimates may be taken as $\hat{\mu}_i = (\Sigma_j y_{ij} x_{ij})/(\Sigma_j y_{ij})$ where x_{ij} is the mid-point of the jth weight-gain interval for the ith diet, and $\hat{\sigma}^2$ as the similarly weighted sample variance. Fitting the model from this starting position provides the estimates $\hat{\mu}_1 = 119{\cdot}1$, $\hat{\mu}_2 = 101{\cdot}5$ and $\hat{\sigma}^2 = 519{\cdot}4$. These compare with the estimates $\hat{\mu}_1 = 120$, $\hat{\mu}_2 = 101$ and $\hat{\sigma}^2 = 446{\cdot}12$ obtained from the original data.

In a manner similar to that for grouped normal data we may treat multiply-classified dose-response data (see, for example, Finney, 1971 pp. 220–226) as Poisson data derived from underlying tolerance distributions, where the expected values are proportional to differences between values of $h(\eta) = \Phi(\eta)$. Most of the regression models for ordinal data of McCullagh (1980) can be treated in the similar way with the same composite matrix but with $h(\eta) = \exp(\eta)/(1 + \exp(\eta))$ or $h(\eta) = -\exp(-\exp(\eta))$.

3.3. *Relative Risks*

The data in Table 1 are taken from Liddell, McDonald and Thomas (1977, Table 7), and refer to dust exposure in lung cancer cases and controls. They were interested in fitting a linear model to the relative risks and here we show how this model might be fitted using the current framework.

We assume that the expected frequencies of the ith control and case groups are p_i and $(\alpha_1 + \alpha_2 x_i)p_i$, where as an example we let x_i be the lower bound of the ith exposure class. The risk of the ith relative to the first class is now

$$[p_i(\alpha_1 + \alpha_2 x_i)/p_i]/[p_1(\alpha_1 + \alpha_2 x_1)/p_1] = 1 + \alpha_2 x_i/\alpha_1 = 1 + \rho x_i$$

since $x_1 = 0$ and $\rho = \alpha_2/\alpha_1$, and so is linear in x_i. This model can now be written in a linear

model form $E(\mathbf{y}) = \boldsymbol{\mu} = Ch(\boldsymbol{\eta})$ with $\mathbf{y}^t = (43\ 10\ 24\ \ldots\ 19\ 8\ 11)^t$. We find

$$C = \begin{bmatrix} I_{10} & O_{10,10} & O_{10,10} \\ O_{10,10} & I_{10} & B_1 \end{bmatrix}, \quad X = \begin{bmatrix} I_{10} & O_{10,2} \\ I_{10} & B_2 \\ I_{10} & B_3 \end{bmatrix},$$

where I_t is an identity matrix of order t while B_1 is a diagonal matrix with ith element x_i,

$$B_2^t = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^t$$

and

$$B_3^t = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}^t.$$

The function $h = \exp$ and $\boldsymbol{\beta}^t = (\ln p_1, \ldots, \ln p_{10}, \ln \alpha_1, \ln \alpha_2)^t$. The observed number of controls may be used as initial estimates of p_i while a linear regression of the ratio cases/controls on x provides initial estimates of α_1 and α_2. Assuming the 20 cell entries are Poisson variables then we obtain $\hat{\rho} = 0{\cdot}00243$ with standard error 0·00073, the fitted values being given in Table 1.

TABLE 1

Dust exposure in lung cancer cases and controls, relative risks and fitted values from a linear relative risk model

Total dust exposure in units of [(million particles per cubic foot) × (years)] up to 7 years before death of the case	*Observed*			*Fitted*		
	Cases	*Controls*	*Relative risk*	*Cases*	*Controls*	*Relative risk*
<6,	43	285	1	43·5	284·5	1·00
$\geqslant 6$, <10	10	62	1·07	9·7	62·3	1·02
$\geqslant 10$, <30	24	166	0·96	25·7	167·3	1·02
$\geqslant 30$, <100	37	211	1·16	35·0	213·0	1·07
$\geqslant 100$, <300	31	168	1·22	31·8	167·2	1·24
$\geqslant 300$, <600	27	95	1·88	25·5	96·5	1·73
$\geqslant 600$, <1000	18	50	2·39	18·6	49·5	2·46
$\geqslant 1000$, <1500	10	19	3·49	10·0	19·0	3·43
$\geqslant 1500$, <2000	6	8	4·97	5·8	8·1	4·65
$\geqslant 2000$	9	11	5·42	9·5	10·5	5·87

4. BILINEAR MODELS

Consideration of the estimation of N_1 and N_2 in the grouped data example suggests an immediate generalization of the composite linear model. We suppose a model where

$$\eta_i = \sum_{j=1}^{p} \beta_j x_{ij}, \quad \gamma_k = h(\eta_k), \quad \eta_{1k} = \sum_{j=1}^{p_1} \beta_{1k} C_k \boldsymbol{\gamma} \quad \text{and} \quad \mu_k = h_1(\eta_{1k}).$$

This is an extension of the composite model in two ways. Firstly, the $n \times m$ matrix C has been replaced by a linear function of parameters β_{1i} i.e. $C = \Sigma \beta_{1j} C_j$, summing over j from 1 to p_1 where C_i are known matrices. Secondly, instead of letting $\mu = \eta_1$ we allow $\boldsymbol{\mu}$ to be some function, h_1, of $\boldsymbol{\eta}_1$. We call this model a generalized bilinear model as two separate linear models are used to specify the model.

The grouped normal data example fits this bilinear framework with $\beta_{11} = N_1$, $\beta_{12} = N_2$,

$$C_1 = \begin{bmatrix} E_1 & O_{9,7} \\ O_{7,9} & O_{7,7} \end{bmatrix}, \quad C_2 = \begin{bmatrix} O_{9,9} & O_{9,7} \\ O_{7,9} & E_2 \end{bmatrix}$$

and h_1 equal to the identity function.

Again Nelder and Wedderburn's argument can be extended to show that the solution of the likelihood equations for β_1 and β is equivalent to an iterative weighted least squares scheme with dependent variate $\boldsymbol{\eta}_1 + CH\boldsymbol{\eta} + H_1(\mathbf{y} - \boldsymbol{\mu})$, independent variates $C_j\boldsymbol{\gamma}$ and CHX where H and H_1 are diagonal $m \times m$ and $n \times n$ matrices with diagonal elements equal to $d\gamma_k/d\eta_k$ and $d\eta_{1k}/d\mu_k$, and with weights $(d\mu_i/d\eta_{1i})^2/\tau_i^2$.

By applying the method of this section to the grouped normal example it can be found that the ML estimates of N_i are given by the sample sizes and that these estimates are uncorrelated with the estimates of the other parameters. In two other cases, truncated grouped normal data and Wadley's problem in bioassay (Finney, 1971, pp. 202–208) bilinear models are appropriate, but the sample sizes N_i are not known. The procedure in this section is then a convenient way of estimating all the parameters.

5. General Composite Link Functions

In Section 2 we discussed the case when $\boldsymbol{\mu}$ is a linear function of $\boldsymbol{\gamma}$. A similar procedure can be used if $\mu_i = c_i(\gamma)$. The working independent and dependent variables are now

$$x_{ij}^* = \sum_{k=1}^{m} (d\mu_i/d\eta_k)\, x_{kj} \quad (j = 1, \ldots, p) \quad \text{and} \quad z_i = \sum_{k=1}^{m} (d\mu_i/d\eta_k) + (y_i - \mu_i)$$

with weight function $1/\tau_i^2$. Also for the bilinear model in Section 4 when

$$\eta_{1k} = \sum_{j=1}^{p_1} \beta_{1j}\, c_{jk}(\gamma) \quad \text{and} \quad \mu_k = h_1(\eta_{1k}),$$

the working independent variables are

$$c_{ji}(\boldsymbol{\gamma}) (j = 1, \ldots, p) \quad \text{and} \quad \sum_{k=1}^{m} (d\eta_{1i}/d\eta_k)\, x_{kj} \, (j = 1, \ldots, p),$$

and the working dependent variate is

$$z_i = \eta_{1i} + \sum_{k=1}^{m} (d\eta_{1i}/d\eta_k)\, \eta_k + (y_i - \mu_i)(d\eta_{1i}/d\mu_i)$$

with weight $(d\mu_i/d\eta_{1i})^2/\tau_i^2$.

6. Efficiency

The composite link functions have been presented because they allow more models to be embedded simply into the generalized linear model framework. Of course in specific cases other procedures might be more efficient computationally. The EM algorithm (Dempster *et al.*, 1977) is a natural competitor. This method, for our models, uses the predicted value $\boldsymbol{\mu} = \mathbf{C}\boldsymbol{\gamma}$ but uses the first differentials of the likelihood to give new estimates. This, we expect, can lead to less computation in each iteration but needs more iterations for convergence. Recently Wolynetz (1979) found this to be true for censored and grouped normal data.

7. Computation

The calculations involved can be easily programmed. It has been found quite simple to perform the calculations using GENSTAT (Nelder *et al.*, 1975) making use of the facilities for matrix operations and fitting linear models. Another possibility is to use GLIM-3 (Baker and Nelder, 1978) which was especially written for fitting generalized linear models.

Using GLIM, the general form of the analysis for any data will be: (1) Input of data, (2) Calculation of values needed in the analysis, (3) Declaration of macros, (4) Declaration of model, (5) Fitting of model, (6) Output of results.

These steps will be described in detail for the example in Section 3.1 (see Baker and Nelder, 1978 for details of the language, etc.):

```
$UNITS 4 $DATA Y $READ 179 6 202 35          $C STEP 1 READ IN
$DATA 9 XP $READ 2 1 1 1 0 0 1 0 0           $C OBSERVATIONS
$DATA 9 XQ $READ 0 1 0 1 2 1 0 1 0           $C AND COLUMNS
$DATA 9 XR $READ 0 0 1 0 0 1 1 1 2           $C OF X
$VARIATE 9 ETA GAM                           $C STEP 2 DECLARE
$VARIATE 36 C $DATA 9 T                      $C TEMPORARY VECTORS
$READ 1 2 4 12 16 23 33 35 36                $C CONSTRUCT C (AS A
$CALCULATE C = 0 : C(T) = 1                  $C VARIATE OF LENGTH 36) AND
$CALCULATE %L = %CU(Y) : %L = %LOG(%L)       $C STORE LOG OF OBSERVATIONS
$MACRO M1 $CAL %A = %NE(%PL,0)               $C STEP 3 USING OWN FACILITY
$SWI%A MEXT                                  $C TO GIVE NON-STANDARD
$CAL ETA = %PE(1)*XP+%PE(2)*XR
     +%PE(3)*XQ                              $C GLM,
$CAL GAM = %EXP(ETA) $CAL %N = 4             $C FOUR MACROS
$WHI %N MAT $ENDMAC                          $C M1,M2,M3,M4
$MACRO M2 $CAL %DR = 1 $ENDMAC               $C ARE NEEDED
$MACRO M3 $CAL %VA = %FV $ENDMAC
$MACRO M4
$CAL %DI = 2*(%YV*%LOG(%YV/%FV)
     +%FV-%YV)
$ENDMAC
$MACRO MEXT $EXTRACT %PE $ENDMAC             $C MACRO M1 USES
$MAC MAT $CAL %N = %N-1 : %I = 4-%N          $C 2 MACROS
$CAL T = C(%GL(9,1)+9*(%I-1))*GAM            $C MEXT AND MAT
$CAL %FV(%I) = %CU(T)
$CAL %LP(%I) = %CU(T*ETA)
$CAL P(%I) = %CU(T*XP)
$CAL Q(%I) = %CU(T*XQ)
$CAL R(%I) = %CU(T*XR) $ENDMAC
$YVA Y $OWN M1 M2 M3 M4 $SCA 1               $C STEP 4 DECLARES MODEL
$VAR 3 %PE $CAL %PE = 1/2*%L-%LOG(3)         $C AND INITIAL VALUES
$CAL %LP = P = Q = R = 0
$FIT P+R+Q-%GM                               $C STEP 5 FITS MODEL
$DISPLAY E V                                 $C 6 OUTPUT OF RESULTS
```

References

Baker, R. J. and Nelder, J. A. (1978). *The GLIM System, Release 3.* Oxford: Numerical Algorithms Group.

Birch, M. W. (1963). Maximum likelihood in three-way contingency tables. *J. R. Statist. Soc.* B, **25**, 230–233.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1971). Maximum likelihood from incomplete data via the EM algorithm (with Discussion). *J. R. Statist. Soc.* B, **39**, 1–38.

Kempthorne, O. (1969). *An Introduction to Genetic Statistics.* Ames, Iowa: Iowa State University Press.

Liddell, F. D. K., McDonald, J. C. and Thomas, D. C. (1977). Methods of cohort analysis: appraisal by application to asbestos mining. *J. R. Statist. Soc.* A, **140**, 469–491.

McCullagh, P. (1980). Regression methods for ordinal data. *J. R. Statist. Soc.* B., **42**, 109–142.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *J. R. Statist. Soc.* A, **135**, 370–383.

Nelder, J. A. *et al.* (1975). Genstat Reference Manual Inter University/Research Council Series Report No. 3, 3rd ed.

Sampford, M. R. (1952). The estimation of response-time distribution. I. Fundamental concepts and general methods. *Biometrics*, **8**, 13–32.

Snedecor, G. W. and Cochran, W. G. (1967). *Statistical Methods*, 6th ed. Iowa: State University Press.

Taylor, G. L. and Prior, A. M. (1938). Blood groups in England. *Ann. Eugen.*, **8**, 343–355.

Wolynetz, M. S. (1979). Algorithm AS 138. Maximum likelihood estimation from confined and censored normal data. *Appl. Statist.*, **28**, 185–195.

ROYAL STATISTICAL SOCIETY
25 Enford Street, London, W1H 2BH

31

A note on the estimation of economic values for selection indices

by

R. Thompson

Anim. Prod. 1980. **31**: 115–117
© 1980. British Society of Animal Production
0003-3561/80/09140115 $02.00

A NOTE ON THE ESTIMATION OF ECONOMIC VALUES FOR SELECTION INDICES

R. THOMPSON

ARC Unit of Statistics, University of Edinburgh, Edinburgh EH9 3JZ

ABSTRACT

A procedure has recently been developed for estimating economic values which employs a profit function implying that animals with mean values for variates in the profit function give the most profit. However, the proposed index selects extreme animals, a paradox which is explained by showing that the index is appropriate for only part of the profit function. More appropriate selection indices are developed and illustrated graphically.

THE PURPOSE of this note is to discuss difficulties in a method of estimating economic values for selection indices suggested recently by Melton, Heady and Willham (1979). Melton *et al.* (1979) argue that, in certain circumstances, profit per animal (¶) may be expressed as

$$¶ = P_y Y - \sum_{i=1}^{n} P_i X_i - FC,$$

where P_y and P_i are the prices per unit of the product and ith input, Y and X_i the amounts of the product and ith input and FC the fixed cost of production. It is assumed that the amount of output Y is a function of the input variables X_i. The input variables are separated into two parts, animal traits ($1 \leqq i \leqq m$) and producer supplied variables ($m < i \leqq n$). Melton *et al.* (1979) are interested in cases where the economic values associated with the animal traits P_i ($1 \leqq i \leqq m$) and the optimum levels of the producer traits X_i ($m < i < \leqq n$) require estimation.

It is suggested that these values, $\hat{P}_j$ ($1 \leqq i \leqq m$) and $\hat{X}_j$ ($m < i \leqq n$) be chosen so that the differentials of the profit function with respect to X_j are zero, i.e.

$$P_y \delta Y / \delta X_i = P_i, \qquad (1)$$

using the argument that the producer wishes to maximize profits. Now $\delta Y / \delta X_i$ in general depends on the values of X_i. Melton *et al.* (1979) suggest evaluating (1) with $X_i = \bar{X}_i$ ($1 \leqq i \leqq m$) to find the economic values and values for the producer traits. This implies that the animal with mean values $\bar{X}_i$ gives the largest estimated profit. Why this should be so is not clear as (1) strictly implies that the slope of the profit function is zero at $X_i = \bar{X}_i$, and this could mean that the profit function was a *minimum* at $X_i = \bar{X}_i$. The point is illustrated below using the example of Melton *et al.* (1979).

Having estimated P_j and X_j, Melton *et al.* (1979) go on to suggest using the common selection index (Hazel, 1943) based on the estimated economic weights $\hat{P}_i$. This is surprising since it will lead to selection of extreme animals on the index scale and we have seen that the 'mean' animal gives the largest estimated profit. The explanation of this paradox is that the index Melton *et al.* (1979) suggest is relevant only if we wish to decrease (or increase!) $\sum_{i=1}^{m} \hat{P}_i X_i$, the estimated input from animal traits. If their derivation of the profit function is accepted it seems more logical to choose an index based on the whole profit function.

In the cases considered by Melton *et al.* (1979) the profit function is a second degree polynomial and a quadratic index can be created using the work of Wilton, Evans and Van Vleck (1968). This is now illustrated using the numerical example of Melton *et al.* (1979), which employs a profit function for beef cattle based on the animal traits average daily gain (ADG) and weaning weight (WW), and the producer trait, days on feed (D). It must be noted that this example, as Melton *et al.* (1979) emphasize, is a simplified demonstration rather than a practical application. Figure 1 plots profit contours against ADG and WW using economic values for ADG and WW, and a value for D given by Melton *et al.* (1979), that approximately satisfies (1). Each

contour shows values of ADG and WW that give rise to the same estimated profit (or, strictly, loss as the profit is less than zero). The profit function is interesting in that the point C, often called a saddlepoint, gives the highest profit in quadrants ACD, BCE but is the lowest value in the quadrants ACE, BCD. The Figure shows that although P_{ADG}, P_{WW} and D are chosen approximately to satisfy (1), and the profit surface is relatively flat in the region of the mean value of ADG (0·785 kg) and WW (193·8 kg), the profit function does not reach its maximum value in this region.

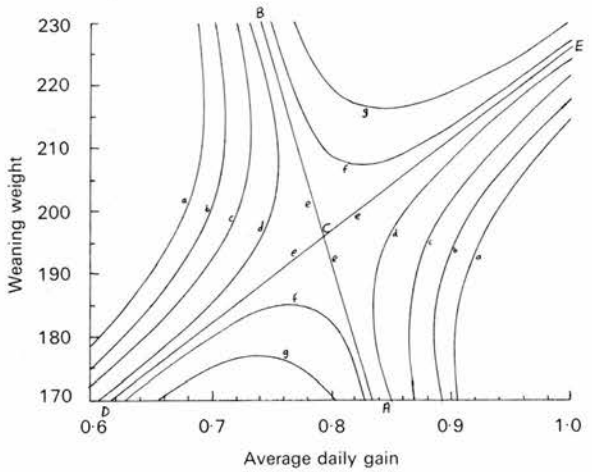


FIG. 1. Profit (¶) contours for the numerical example of Melton *et al.* (1979). ¶ = −$184 (a), ¶ = −$184.2 (b), ¶ = −$184·4 (c), ¶ = −$184·6 (d), ¶ = −$184·721 (e), ¶ = −$184·8 (f), ¶ = −$185·0 (g).

Wilton *et al.* (1968) give a quadratic index, I, that maximizes the correlation between ¶ and I when ¶ is a second order polynomial. It is convenient to construct this index in two stages.

First, predictions of ADG(WW) are found by using linear selection index theory with economic weights 1(0) for ADG and 0(1) for WW. For example,

$$B_1\ (10\,000) + B_2(320) = 1(2600) + 0(424) \qquad (2)$$
$$B_1\quad (320) + B_2(344) = 1(424)\ + 0(103)$$

(compare with equation (22) of Melton *et al.*, 1979). From (2), $B_1 = 0{\cdot}227$ and $B_2 = 0{\cdot}00102$ and gives rise to the equation

$$\Delta ADG^{\star} = 0{\cdot}277\ \Delta ADG + 0{\cdot}00102\ \Delta WW, \quad (3)$$

and similarly

$$\Delta WW^{\star} = 33{\cdot}826\ \Delta ADG + 0{\cdot}2680\ \Delta WW \quad (4)$$

where Δ indicates deviation from mean value and $\star$ indicates predicted value. Substituting the predicted values from (3) and (4) into the profit function of Melton *et al.* (1979) (equation (19)) gives

$$\begin{aligned} I = {} & 103{\cdot}8071 + 17{\cdot}3624\,\Delta ADG - 0{\cdot}15039\,\Delta WW \\ & + 0{\cdot}11721\,D \\ & - 0{\cdot}05861\,\Delta ADG^2 - 0{\cdot}0000493\,\Delta WW^2 \\ & - 0{\cdot}00282\,D^2 \\ & + 0{\cdot}12042\,\Delta ADG \times D \\ & + 0{\cdot}0051886\,\Delta WW \times D \\ & - 0{\cdot}0098976\,\Delta ADG \times \Delta WW \\ & - 0{\cdot}785\,P_{ADG} - 193{\cdot}8\,P_{WW} \\ & - 0{\cdot}22732\,P_{ADG} \times \Delta ADG \\ & - 33{\cdot}82550\,P_{WW} \times \Delta ADG - 0{\cdot}0010211\,P_{ADG} \\ & \times \Delta WW - 0{\cdot}26745\,P_{WW} \times \Delta WW. \end{aligned}$$

Figure 2 gives contours of I against ADG and WW. The predicted genetic merit surface is the same shape as the profit surface in Figure 1 but is flatter and with a different orientation. Again, C represents the highest value of I in quadrants

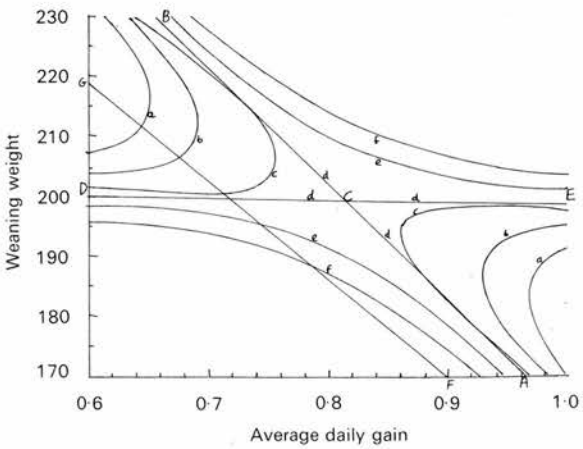


FIG. 2. Quadratic index (I) contours from equation 3. I = −$184·71 (a), I = −$184·715 (b), I = −$184·72 (c), I = −$184·721 (d), I = −$184·725 (e), I = −$184·73 (f). Contours of the linear index of Melton *et al.* (1979) are parallel to FG.

ACD and BCE but the lowest value in quadrants ACE and BCD. The linear index suggested by Melton *et al.* (1979) gives rise to contours parallel to FG in Figure 2. Extreme (either high or low) animals on this linear index are low animals on the quadratic index.

This quadratic index based on the genetic merit of individuals has been introduced simply to show the limitations of the approach of Melton *et al.* (1979). However, it is questionable whether this is the most appropriate index. If one is interested in the improvement of offspring, then a similar index could be developed that gives different weight to the linear and quadratic parts. An alternative approach by Moav and Hill (1966) based on improving the offspring of the 'average' parents uses graphical techniques to derive linear indices for non-linear profit functions.

Quadratic functions have been suggested for estimating aggregate genotype in cases where optima are desired at intermediate, rather than extreme values. The dangers of not inspecting the resulting surfaces are indicated quite startlingly by this example. Even when a surface is known to have an intermediate optimum, then a quadratic surface fitted to it might be similar to Figure 1. Sanders (1977) has an example of this type when predicting efficiency in beef cattle in terms of weaning weight and 3-year weight. Of course, if a profit surface similar to Figure 1 is correct then there are two different strategies for improvement, i.e. using animals in quadrant ACB or in quadrant DCE.

REFERENCES

HAZEL, L. N. 1943. The genetic basis for construction of selection indices. *Genetics, Princeton* **28:** 476–490.

MELTON, B. E., HEADY, E. O. and WILLHAM, R. L. 1979. Estimation of economic values for selection indices. *Anim. Prod.* **28:** 279–286.

MOAV, R. and HILL, W. G. 1966. Specialised sire and dam lines. IV. Selection within lines. *Anim. Prod.* **8:** 375–390.

SANDERS, J. D. 1977. Application of a beef cattle production model to the evaluation of genetic selection criteria. *Ph.D. Thesis, Dep. Anim. Sci., Texas A. and M. Univ.*

WILTON, J. W., EVANS, D. A. and VAN VLECK, L. D. 1968. Selection indices for quadratic models of total merit. *Biometrics* **24:** 937–949.

(*Received* 16 *January* 1980—*Accepted* 6 *March* 1980)

32

Genetic and environmental influences on beef cattle production in Zambia. 1. Factors affecting weaner production from Angoni, Barotse and Boran dams

by

W. Thorpe, D.K.R. Cruickshank and R. Thompson

Anim. Prod. 1980, **30**: 217–234

GENETIC AND ENVIRONMENTAL INFLUENCES ON BEEF CATTLE PRODUCTION IN ZAMBIA

1. FACTORS AFFECTING WEANER PRODUCTION FROM ANGONI, BAROTSE AND BORAN DAMS

W. THORPE AND D. K. R. CRUICKSHANK

Central Research Station, Mazabuka, Zambia

AND

R. THOMPSON

ARC Unit of Statistics, University of Edinburgh, Edinburgh EH9 3JZ

ABSTRACT

Factors influencing reproductive performance and calf and dam live weights were examined for two indigenous breeds, Angoni and Barotse, and an introduced breed, Boran, under ranching conditions in Zambia.

Over 5 years, 675 Angoni, 731 Barotse and 815 Boran records gave calving percentages of 82·5, 78·1 and 75·4 respectively. Year was an important influence on calving and weaning percentages in all breeds; age of dam had little effect and status at mating was only important in the Angoni breed, in which lactating dams produced more calves than dry dams.

Mean calf weaning weights were 147·3, 167·0 and 169·5 kg for the Angoni, Barotse and Boran dam breeds respectively. Hereford, Friesian and Boran bulls sired heavier progeny than Angoni and Barotse. Dam age and status did not influence calf weights, which were positively correlated with dam weights and negatively correlated with dam weight change during lactation.

Dam live weights at the beginning of the breeding season were 313, 357 and 348 kg for the Angoni, Barotse and Boran breeds respectively. Angoni and Boran dams reached maximum live weight at 7 years and Barotses at 9 years. Subsequent calving rate was not influenced by live weight at the beginning of the breeding season nor, in lactating dams, by live-weight change between the autumn peak and mid breeding season.

INTRODUCTION

IN 1965 the Department of Agriculture began a programme of beef cattle research at the Central Research Station, Mazabuka, Zambia. The main aim was to evaluate under ranching conditions the beef production of most of the major breeds available. The Angoni, Barotse and Boran breeds were evaluated both as sire and dam breeds, the Friesian and Hereford as sire breeds

and the Africander as a pure breed. Wherever possible it was planned to evaluate the purebred and crossbred genotypes for reproductive performance, growth and carcass production.

The Angoni breed, indigenous to the Eastern Province of Zambia, and the Barotse breed, indigenous to the Western Province, constitute about 15% and 25% respectively of the national herd of some 1·8 million. These breeds are found mainly in the large traditional sector in which grazing is communal on unenclosed land, and management practices are few and of a low standard. Traditionally managed cattle comprise about 85% of the national herd. The remainder is found in the commercial sector on fenced natural grassland ranches and management includes the retention of standing hay for dry-season grazing, a restricted breeding season, the separation of breeding and young stock and regular disease control measures. The introduced Africander and Boran breeds, which originate from south and east Africa respectively, are found primarily in the commercial sector.

Mason and Maule (1960) have classified the Angoni and Boran breeds as short-horned Zebus, and the Africander and Barotse breeds as long-horned Sangas. The Africander and Boran are often mated to the *Bos taurus* breeds, Hereford and Sussex, on the commercial ranches in Zambia. The *Bos taurus* beef populations are based on imported stock from Britain and South Africa and some artificial insemination with British semen is practised. The small commercial dairy industry is based on Friesians imported as live cattle or semen mainly from Europe, Kenya and South Africa. Friesian × indigenous breed females are distributed as part of a government scheme encouraging milk production in the rural areas.

In anticipation of the continued development of a beef industry, the research programme began with foundation breeding herds of purebred Angoni, Barotse and Boran dams. In this paper analyses of records from these herds provide estimates of factors affecting calving and weaning percentages, calf birth and weaning weights, and dam live weights within each of these breeds. In addition, estimates of the relationships between calf and dam live weights and between dam reproductive performance and dam live weights and live-weight change are reported. Subsequent papers will report aspects of purebred and crossbred progeny performance, including estimates of heterosis and maternal effects, for live weights from birth to slaughter, male castrate carcass traits and female reproductive performance.

MATERIAL AND METHODS

Breeding programme. At the beginning of the programme purebred Angoni, Barotse and Boran breeding females of known age and previous parous state were available at the Central Research Station, Mazabuka. Results are reported for the five breeding seasons, 1966 to 1970. In all years purebred Angoni, Barotse and Boran progeny were produced. In addition, in 1967 and 1969 the six two-way crosses between these breeds were produced, and in 1969 and 1970 Friesian and Hereford adult male cattle (bulls) sired progeny from the three dam breeds. Most of the purebred male progeny born in 1966, 1968 and 1970 were left entire. Thirteen Angoni, 12 Barotse, 13 Boran, 5 Friesian and 8 Hereford bulls sired progeny during the programme.

Environment. All the records were collected at the Central Research Station, Mazabuka, latitude 15°50′S and longitude 27°45′E, which is situated

on the edge of the peneplain of the Kafue River at an altitude of 990 m. The station lies within the 24°C isotherm with absolute temperatures of 39·4°C maximum and 3·3°C minimum, and a mean annual temperature of 22·0°C. The dry season, May to September, is coolest from June to August and temperatures rise to a maximum in October. Effective rain falls between November and March with large annual and monthly variation. Mean annual rainfall for the period of the study was 715 mm.

Some 3500 ha of natural upland veld were available for year-round grazing and an adjoining area of river flats was used by the breeding females for grazing in the latter half of the dry season. In Zambia, where most grazing is unimproved veld, seasonal variation in the growth and quality of herbage closely follows the annual rainfall pattern. This results in a fluctuating nutritional level which gives rapid weight gains during the rains and maintenance or submaintenance nutrition in the dry season. The cattle grazing the river flats during the dry season had the benefit of utilizing fresh herbage growth.

Management. The breeding season was limited to a 3-month period from late December to early March. The ample good-quality grazing available during this period provided improving nutritional conditions for milk production and reconception. Breeding was by natural service, mainly in single-sire herds. After the breeding season cows and calves were run in breed herds under comparable conditions. Male calves were castrated in April; all calves were weaned at an average age of 7·5 months over a 3-day period in early June.

Young adult females (heifers) were introduced into the breeding herds for mating when they were 2 years of age in 1966 and 1967 and at 3 years thereafter. All heifers except those with low live weights were added to the breeding herds. Dams failing to calve in two consecutive seasons had subsequent records omitted from the analyses. Breeding females and their calves did not receive supplementary feed.

While on the upland grazing all cattle were sprayed weekly with an acaricide and weighed monthly and at weaning. Cattle were not weighed or sprayed while grazing on the flats. The breeding females were moved from the upland to the river flats in August and calving began in late September. Calves were recorded within 24 h of birth. Cows and calves were brought back to the upland grazing before the rainy season was established and calving was completed by mid-December. As the flats were flooded during the dry season of 1969 the breeding herds remained on the upland grazing for that calving season.

Description of data and statistical analyses. Dam records were built up from progeny records collected between 1965 and 1970 and included the dam's age and date of last calving, her live weights at the beginning and end of the breeding season during which the calf was conceived, and the live weights at the autumn peak and on the day the previous calf crop was measured. As dam live weights were not routinely recorded at the beginning and end of the breeding season the monthly live weights nearest to the appropriate dates were used. The live weight recorded in the month in which most dams reached their maximum live weight was assumed to be the autumn peak.

A small number of records, 4·0, 2·6 and 0·5% for the Angoni, Barotse and Boran dams respectively, were excluded from the analyses. The major reason

for this was out-of-season calving. Table 1 gives the number of dam records classified by the major factors fitted in the analyses.

Analyses were carried out for each dam breed separately to provide estimates of environmental effects. Estimates of heterosis and maternal effects are reported in the later papers in this series. The traits analysed were calving and weaning percentages (calculated as the proportion of live calves

TABLE 1

Number of dam records classified by breed, year, sire breed, dam age and dam status

Dam breed	Angoni	Barotse	Boran
Year			
1966	95	120	138
1967	110	134	147
1968	123	142	150
1969	164	169	180
1970	183	166	199
Sire breed			
Angoni	432	89	92
Barotse	78	472	91
Boran	81	86	523
Friesian	27	58	54
Hereford	57	26	54
Dam age at calving (years)			
3	24	21	25
4	138	116	149
5	112	105	112
6	76	82	98
7	59	74	90
8	56	83	86
9	54	72	71
10	49	58	66
11	43	46	50
12	32	38	36
$\geqslant 13$	32	36	31
Dam status at mating			
Dry	201	196	258
Lactating, 3 and 4 years	29	36	33
Lactating, $\geqslant 5$ years			
early calved	262	234	262
mid calved	110	162	171
late calved	73	103	90
Total	675	731	814

born and weaned to dams mated each season), calf birth and weaning weights, and the dam live weights mentioned above.

In the hierarchical analyses of variance for calving and weaning percentages effects were fitted for year (1966 to 1970), sire breed (Angoni, Barotse, Boran, Friesian, Hereford), dam age at calving (3, 4 . . . 12, $\geqslant 13$) and dam status at mating (dry; $\leqslant 4$ years lactating; $\geqslant 5$ years lactating < 294 days, 294 to 314 days, > 314 days). All lactating dams, $\geqslant 5$ years, were subdivided according to calving date to estimate the confounded effects of

nutrition during the early months of lactation, and the time interval between calving and the beginning of the breeding season. The calving data classes grouped calves born before 21 October (< 294 days), between 22 October and 10 November (294 to 314 days) and after 10 November (> 314 days). The rainy season was generally established by day 315. Dams which were not lactating at the beginning of the breeding season were included in the dam status class 'dry'. The 'dry' class included dams which were exposed to the bull in the previous breeding season but which did not calve, and dams which were exposed to the bull for the first time (heifers). The dams status class, ⩽4 years lactating, includes those 'heifers' which calved in their first season and hence were lactating when exposed to the bull for the second time. For the calf weights, effects were fitted for year, sire breed, dam age at calving, dam status at mating (dry, ⩽4 years lactating, ⩾5 years lactating), sex (entire male, castrated male, female), period of birth (< 294 days, 294 to 314 days, > 314 days) and the covariate date of birth. The model for dam live weights was the same as that for calving percentage with the addition of the effects of dam's previous parous state (dry, ⩽4 years lactating, ⩾5 years lactating) and dam status at next calving (calved, not calved). Further analyses derived the relationships between weight at the beginning of the breeding season and conception rate as measured by subsequent calving percentage. The relationships were estimated by analysing calving percentage within the same model as used previously with the addition of the effects of dam's previous parous state and live weight at beginning of the breeding season grouped in 10 kg intervals. A similar approach was taken when estimating, within the lactating population, the relationship between dam live-weight change from the autumn peak to mid breeding season and subsequent calving percentage. The live-weight change was expressed as a percentage of dam autumn peak weight and grouped in 3% intervals of the dam's autumn peak weight.

RESULTS

Calving percentage. Mean calving percentages were 82·5, 78·1 and 75·4 for the Angoni, Barotse and Boran dam breeds respectively. The estimates for the effects of sire breed, year of calving, dam age at calving and dam status at mating are given in Table 2.

Sire breeds did not generally have a significant effect on calving percentage, although for the Angoni and Barotse dams higher percentages were achieved when mated to sires of their own breed. For the Boran dams the Angoni sires gave the highest calving percentage. The small number of records for the *Bos taurus* sire breeds gave very variable results. Year effects were consistent in that calving rate was significantly lower in 1970 than in 1966 in each of the three breeds, and there was a general decline in reproductive performance between these years. The Barotse performance was particularly depressed in 1970.

The effect of dam status at mating differed in the three breeds. For the Barotses the results showed that dry dams (which included heifers) tended to calve at a higher rate than the lactating dams. By contrast, the lactating Angoni and Boran dams had higher calving percentages than dry dams. The differences were significant for the Angoni breed in which the older dry cows had a particularly low calving percentage. For the Angoni and Barotse breeds, but not the Boran, the earlier a calf was born during the calving season

TABLE 2

Contrasts for main effects on calving percentage in dams of the Angoni, Barotse and Boran breeds

Dam breed	Angoni	Barotse	Boran
Overall mean	82·5	78·1	75·4
Sire breeds			
Differences from Angoni			
Barotse	−4·4	3·7	−3·3
Boran	−7·4	1·9	−3·5
Friesian	−1·7	−1·9	−19·3
Hereford	−11·7	13·8	−17·1
Av. s.e. of differences	7·42	7·74	7·11
Year			
Differences from 1966			
1967	−10·4	10·1	−2·5
1968	−11·5	−0·9	−10·4
1969	−15·8	−7·2	−8·9
1970	−23·0	−28·2	−14·4
Av. s.e. of differences	5·04	5·31	5·19
Dam age at calving (years)			
Differences from 3 years			
4	8·6	9·0	8·5
5	−6·7	22·1	11·9
6	−7·4	21·0	11·5
7	−5·9	25·4	21·5
8	−2·0	19·5	18·4
9	−9·2	14·4	18·8
10	−2·8	17·9	16·9
11	−2·4	16·1	5·5
12	−1·1	15·7	24·1
⩾13	−14·8	10·8	8·2
Av. s.e. of differences	7·96	8·06	8·12
Dam status at mating			
Differences from dry dams			
Lactating, 3 and 4 years	2·2	−4·3	11·2
Lactating, ⩾5 years			
early calved	17·5	5·4	4·4
mid calved	12·8	−6·6	0·4
late calved	10·2	−10·6	4·1
Av. s.e. of differences	7·18	6·77	7·15

the more likely was its dam to reconceive. Differences between the dam status classes were not significant in the Barotse and Boran breeds.

Age had a significant effect on calving percentage in the Angoni breed, in which 4-year-old dams calved at a higher rate than dams of most older ages. By contrast, the calving rates of Barotse dams aged 3 and 4 years were significantly lower than the rates of 5-, 6- and 7-year-old dams, and 3-year-old Boran dams had a significantly lower calving rate than some older age classes. Other differences in each of the breeds were not significant. These results must be qualified. This is because the dry dam status class included all 3- and some 4-year-old dams; and in the Angoni, but not in the Barotse and Boran breeds, dry dams calved at a significantly lower rate than the lactating dams

of ⩾5 years. Adjustment of the age estimates for these dam status effects will have altered the relative estimates in the Angoni breed which, when age effects were considered without reference to dam status, showed no significant differences between the age classes.

Weaning percentage and preweaning mortality. Mean weaning percentages for the Angoni, Barotse and Boran dam breeds were 80·3, 74·0 and 69·2 respectively. The influence of year of birth, dam age at calving and dam status at mating were little different from those described for calving percentage.

Differences between calving and weaning percentages gave estimates of pre-weaning mortality. Although no formal analyses were carried out it was apparent that breed differences were large, with percentage mortalities of 2·7, 5·3 and 8·3 for the Angoni, Barotse and Boran dam breeds respectively. Sire breed percentage means were very similar to the dam breed means, 2·5, 5·0 and 9·4 respectively. Records for the *Bos taurus* sire breeds were too few to give valid comparisons.

Calf birth and weaning weights. Table 3 presents the effects of sire breed,

TABLE 3

Contrasts for sire breed, time and sex effects on calf birth and weaning weights (kg) in dams of the Angoni, Barotse and Boran breeds

	Birth weight			Weaning weight		
Dam breed	Angoni	Barotse	Boran	Angoni	Barotse	Boran
Overall mean	22·9	25·7	25·2	147·3	167·0	169·5
Sire breed						
Difference from Angoni						
Barotse	0·4	0·2	0·4	3·1	2·8	9·6
Boran	1·8	3·1	1·3	7·1	13·9	9·7
Friesian	4·2	5·1	3·6	25·5	23·4	26·0
Hereford	3·2	2·5	2·5	24·9	26·9	32·2
Av. s.e. of differences	0·71	0·60	0·71	4·26	3·95	4·01
Year						
Differences from 1966						
1967	−0·1	−0·7	−0·0	7·4	3·5	14·0
1968	−0·5	−0·7	−1·4	11·2	2·4	6·0
1969	0·2	−0·5	0·3	−1·0	−9·6	4·9
1970	1·0	1·2	2·2	0·1	−2·9	6·8
Av. s.e. of differences	0·44	0·45	0·48	2·62	2·66	2·79
Period of birth						
Differences from <294 days						
294–314 days	1·1	0·1	0·6	0·7	1·9	−1·1
⩾315 days	−0·4	−0·1	0·9	−8·7	−1·6	−3·7
Av. s.e. of differences	0·65	0·67	0·70	3·92	3·47	3·86
Date of birth	0·028	0·032	0·008	−0·358	−0·575	−0·520
s.e.	0·017	0·018	0·020	0·101	0·101	0·114
Sex						
Differences from female						
Entire male	1·7	1·8	1·7	14·5	17·3	18·5
Male castrate	—	—	—	16·5	15·4	11·1
Av. s.e. of differences	0·28	0·27	0·28	2·64	2·26	2·42

year, period and date of birth, and sex on calf birth and weaning weights, which averaged 22·9, 25·7 and 25·2 kg at birth, and 147·3, 167·0 and 169·5 kg at weaning, for the Angoni, Barotse and Boran dam breeds respectively.

Dams of the three breeds produced heavier calves at birth when mated to Friesian, Hereford and Boran bulls than when mated to Barotse and Angoni bulls. Generally these differences were statistically significant, as were the birth weight advantages of the Friesian- and Hereford-sired calves over the Boran-sired calves. The differences between the Angoni and Barotse and between the Friesian and Hereford sire breeds were small and non-significant. The sire breed birth weight advantages were generally maintained to weaning, except that the Boran-sired progeny were only significantly superior to the Barotse-sired progeny when born to Barotse dams.

Year of birth had a significant effect upon calf weights. Calves born in 1970 were significantly heavier at birth than calves born in all other years in each of the three breeds. The lightest calves at birth in all breeds were born in 1968, and the Boran birth weight in that year was significantly lighter than in any other year. The effect of year of birth on calf weaning weight in the three breeds was much more variable.

The birth and weaning weights of calves born to Barotse and Boran dams were not influenced by period of birth (Table 3). On the other hand, calves born to Angoni dams during the middle of the calving season had significantly heavier birth weights than early-born calves. At weaning the late-born calves of Angoni dams were significantly lighter than early-born calves, and lighter than the weaners born during the middle period. Calves born in the last period were also the lightest at weaning in the Barotse and Boran breeds. The covariate, date of birth, was not significant for birth weight in any of the dam breeds, but at weaning had a significant effect in all breeds, such that calves born 30 days later than average were 11, 17 and 17 kg lighter at weaning in the Angoni, Barotse and Boran dam breeds respectively. For all three breeds male calves were significantly heavier than castrated males at weaning in the Boran breed.

The effects of dam age at calving and dam status at mating on birth weight were small (Table 4). In general, dams lactating at mating gave birth to lighter calves than dams which were dry, and calves born to dams of intermediate age tended to be heavier than calves born to young and old dams. The only significant difference was that between the heavier birth weights of calves born to Barotse dams lactating at mating, as opposed to calves born to Barotse dams dry at mating. While dam status had no significant effect on calf weaning weights, dams calving in successive years generally weaned lighter calves than dams which had not calved the previous year. Dams 9 years of age at calving gave the heaviest weaning weights for the Barotse breed, weights which were significantly heavier than calves weaned by dams which were 4 years of age. There were no significant differences between the weaning weights of the dam age classes in the Angoni and Boran breeds.

Dam live weights. Mean live weights at the beginning and end of the breeding season, at the autumn peak, and at weaning were 313·0, 338·2, 344·5 kg; 333·1, 357·0, 386·9 kg; 392·7, 378·7, 348·1 kg; and 378·8, 384·6, 375·3 kg for the Angoni, Barotse and Boran dam breeds respectively. Dams gained weight during the breeding season, which coincided with the rains, and lost weight between the autumn peak and weaning.

Year effects on dam live weights were large in each of the breeds, with the

TABLE 4

Contrasts for dam age at calving and dam status at mating effects on calf birth and weaning weights (kg) in dams of the Angoni, Barotse and Boran breeds

	Birth weight			Weaning weight		
Dam breed	Angoni	Barotse	Boran	Angoni	Barotse	Boran
Overall mean	22·9	25·7	25·2	147·3	167·0	169·5
Dam age at calving (years)						
Differences from 3 years						
4	−0·1	−0·1	0·3	4·0	−3·7	−2·5
5	0·6	0·1	−0·2	5·6	1·8	−6·9
6	0·5	0·3	−0·1	7·8	3·8	−8·5
7	0·1	0·7	0·4	8·8	5·9	−4·6
8	−0·1	0·6	0·3	10·4	8·4	−5·4
9	0·4	0·6	0·5	9·5	9·1	1·1
10	−0·1	0·5	0·1	10·9	6·7	−5·0
11	−0·3	−0·3	0·1	6·2	5·5	−3·1
12	−1·0	−0·4	−0·5	7·7	9·3	−7·1
⩾13	−0·6	−0·5	−0·5	9·4	9·3	−7·6
Av. s.e. of differences	0·73	0·69	0·76	4·31	3·83	4·23
Dam status at mating						
Differences from dry dams						
Lactating, 3 and 4 years	0·5	1·9	−0·7	−0·1	−5·2	−7·7
Lactating, ⩾5 years	−0·3	1·0	−0·3	−4·6	−4·2	1·8
Av. s.e. of differences	0·70	0·66	0·70	4·24	3·81	3·96

weights prior to weaning generally significantly lower in 1970 than in other years. Dam weaning weights were lower in both 1968 and in 1970 than in other years in all breeds, and weaning weight of the Barotse dams in 1969 was relatively lower than those of the Angoni and Boran dams.

Figure 1 illustrates the effect of dam age on live weight at the beginning of the breeding season. While the effect of age was similar up to 7 years of age in each of the breeds, thereafter there was no significant increase in live weight except in the Barotse breed, which at the beginning of the breeding season reached maximum weight at 9 years of age. Results for the three later weights gave very similar results.

Dam status at mating and the dam's previous parous state gave significant differences for weight at the beginning of the breeding season. Dams dry at mating were significantly heavier than lactating dams of ⩾5 years in all breeds: for example, the early calving groups were 27·1, 20·1 and 31·2 kg lighter than dry dams in the Angoni, Barotse and Boran breeds respectively. The late calvers usually had significantly heavier weights at the beginning of the breeding season than early calvers. By weaning, these differences had disappeared in the Angoni and Barotse breeds but not in the Boran breed. Dam weight at the beginning of the breeding season was also affected by previous parous state; dams which calved the previous year were significantly lighter than previously barren dams. For ⩾5 year dams the differences were −13·2, −18·7 and −16·5 kg for Angoni, Barotse and Boran dams respectively. Previously barren Barotse dams were still significantly heavier at weaning, but in the other breeds the differences were smaller and non-significant.

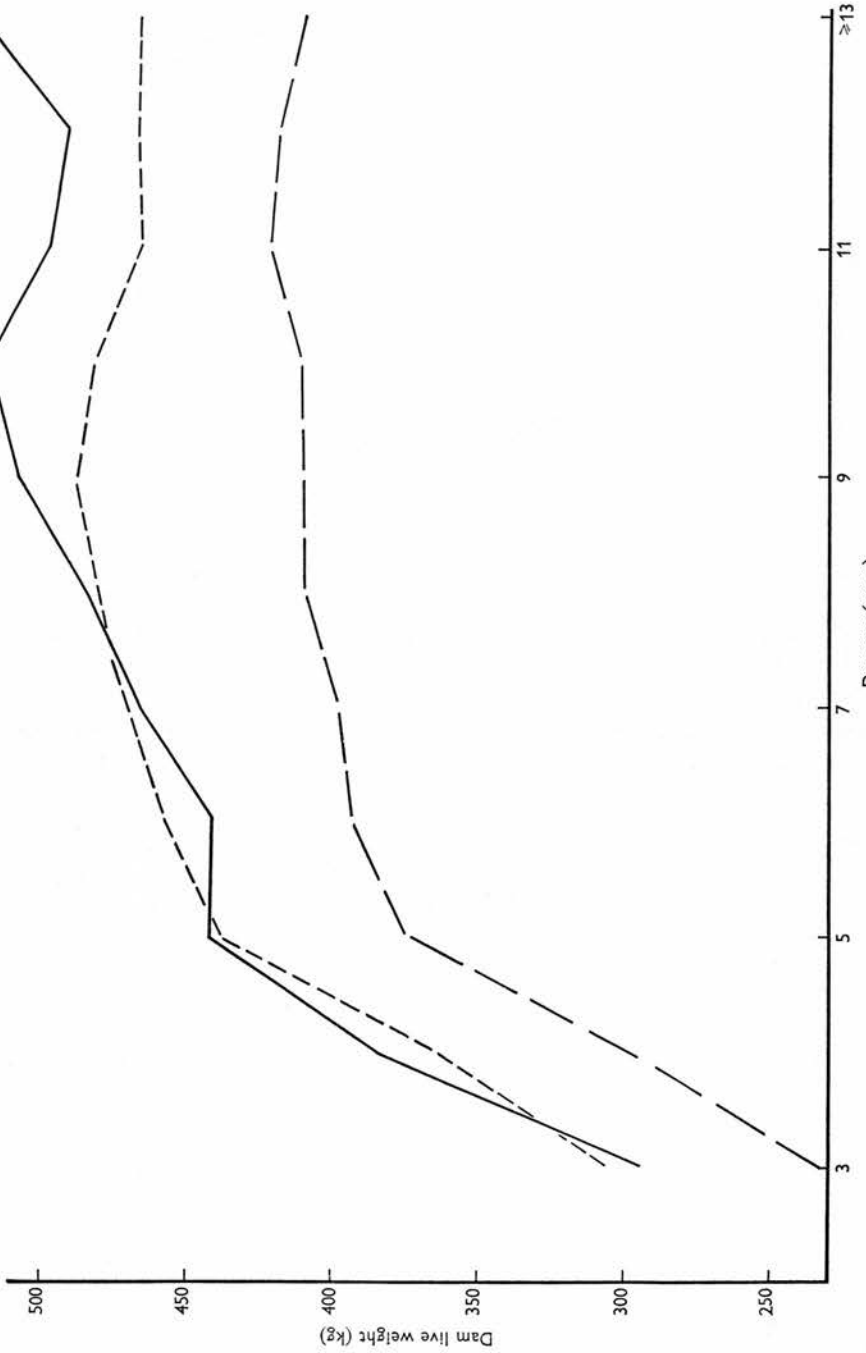


FIG. 1. Effect of dam age on live weight at the beginning of the breeding season in the Angoni, Barotse and Boran breeds (——— Barotse; — — — — — Boran; ——— ——— Angoni).

Relationships between calving percentage and dam live weights and live-weight changes. In the dam live weight breed analyses described above, a covariate estimated the influence of dam live weight on conception rate as measured by calving rate in the following season. Dams which subsequently calved were marginally heavier at the beginning and end of the breeding season, and at weaning, than dams which did not calve. The differences were smallest in the Angoni breed and intermediate in the Boran. The differences approached significance in the Barotse breed, and increased from +5·7 kg at the beginning of the breeding season to +14·7 kg at weaning. Weaning occurred a minimum of 3 months and a maximum of 6 months after conception.

The other approach taken was to consider the effect of live weight on subsequent calving percentage. Figure 2 shows that there were no significant relationships between subsequent calving percentage and dam weight at the beginning of the breeding season grouped by 10 kg intervals, although there was some indication that dams in the heavier live weight groups had higher calving percentages. Boran dam results were particularly erratic. Similarly inconsistent and inconclusive results were found when live-weight changes during the individual breeding seasons were related to subsequent calving percentages.

It was possible to examine the relationship between the subsequent calving percentage of lactating dams, and their percentage live-weight change from the autumn peak to the mean of the live weights at the beginning and end of the breeding season in the latter 4 years of the data. Mean percentage live-weight changes were very similar in the Angoni, Barotse and Boran breeds,

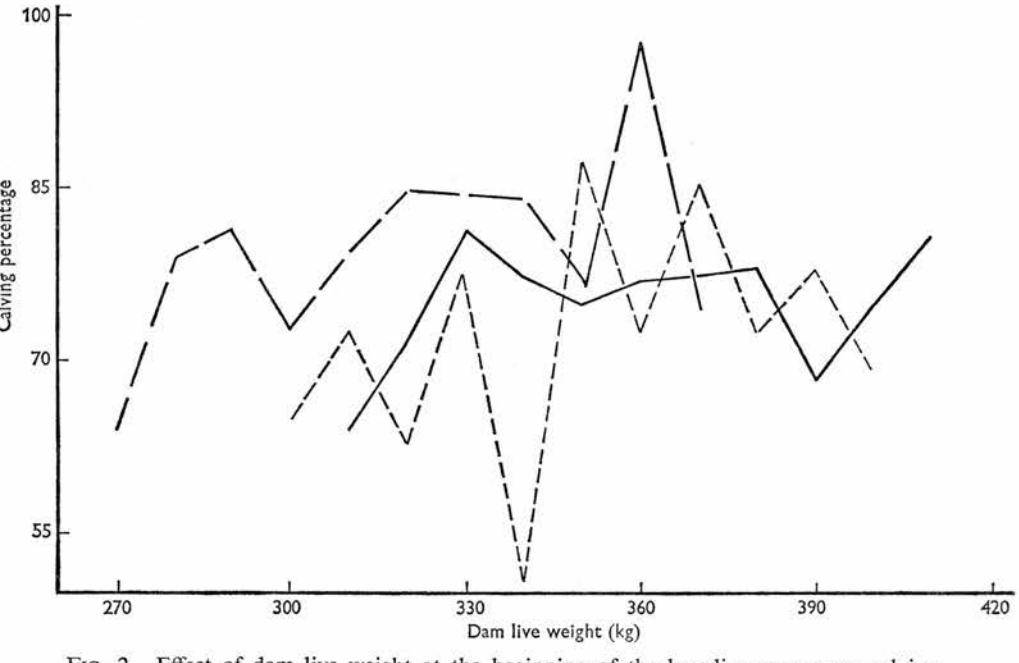


FIG. 2. Effect of dam live weight at the beginning of the breeding season on calving percentage in the Angoni, Barotse and Boran breeds (———— Barotse; — — — — — — Boran; ——— ——— Angoni).

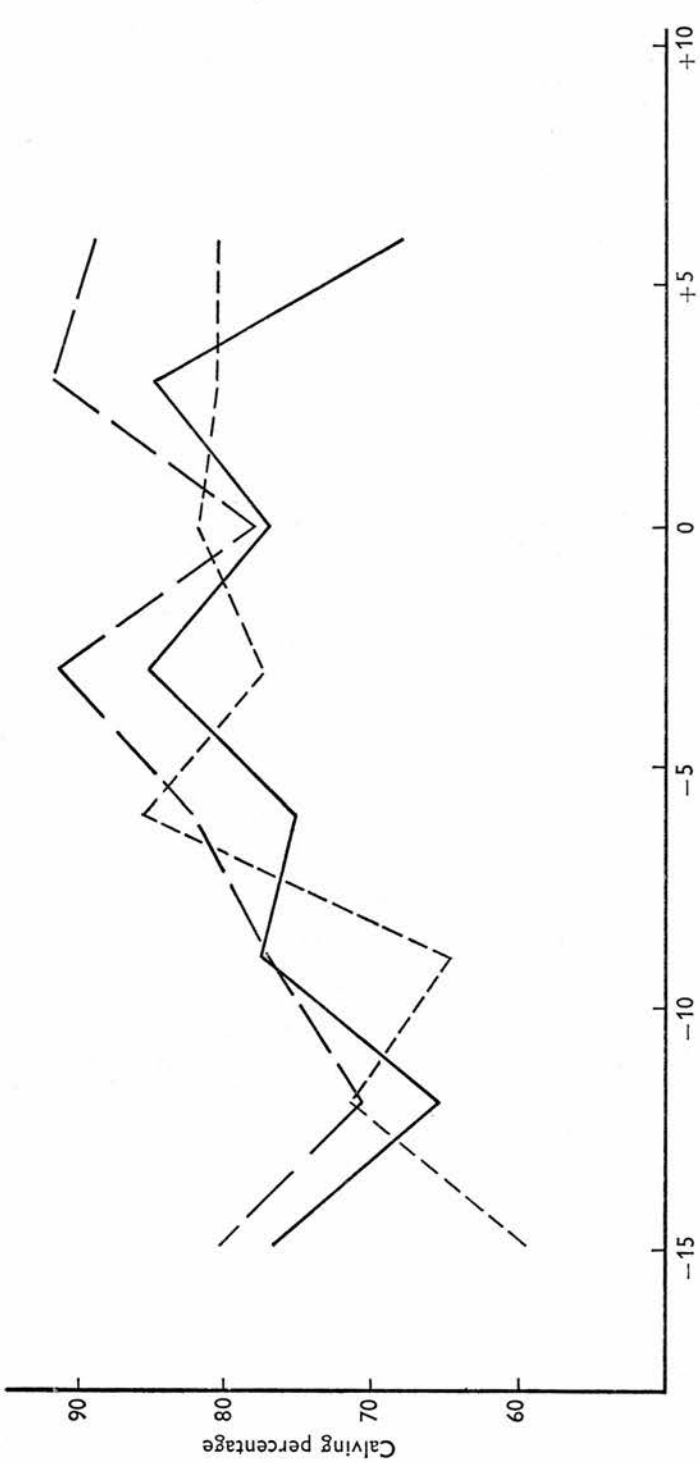


FIG. 3. Relationship between subsequent calving percentage and percentage live-weight change from autumn peak to mid mating for dams of the Angoni, Barotse and Boran breeds (———— Barotse; — — — — — Boran; ——— ——— Angoni).

−6·6, −6·0 and −6·6% respectively. Figure 3 shows the subsequent calving percentage of dams grouped by 3% live-weight change intervals. Again the results were inconsistent although there is some indication, particularly in the Boran, that higher calving percentages were achieved by those dams losing no more than 5% live weight between the autumn peak and mid breeding season weights. The regression coefficients describing the relationship were 0·0054 ± 0·0045, 0·0024 ± 0·0045 and 0·0074 ± 0·0042 for the Angoni, Barotse and Boran dams respectively. The regressions indicate that for each 10% weight loss within the range measured here, calving rate was depressed by about 15, 11 and 5% for the Boran, Angoni and Barotse dams respectively. The large standard errors of the regression coefficients show that these estimates have low predictive value.

Relationships between dam live weights and calf birth and weaning weights. Table 5 shows the correlations and regressions for calf and dam live weights

TABLE 5

Correlations and regressions for calf and dam live weights and live-weight changes

	Measures related							
	BW and BBS†		WW and BBS		WW and DWW		WW and DCW	
Dam breed	*r*	*b*	*r*	*b*	*r*	*b*	*r*	*b*
Angoni	0·192	0·021	0·172	0·112	0·074	0·044	−0·165	−0·195
Barotse	0·256	0·023	0·212	0·101	0·156	0·072	−0·101	−0·096
Boran	0·208	0·021	0·222	0·132	0·134	0·078	−0·163	−0·188
Av. s.e.	0·0479	0·015	0·0481	0·032	0·0485	0·025	0·0485	0·040

† In this and Table 6, BW: calf birth weight; WW: calf weaning weight; BBS: dam weight at the beginning of the breeding season; DWW: dam weight at weaning; DCW: dam change in weight between BBS and DWW.

and live-weight change estimated from the records of weaned calves and their dams. The correlations of calf birth and weaning weights with their dam's weight recorded at the beginning of the breeding season, when progeny mean age was 2·5 months, were about 0·20. These estimates are the pooled sums of the between-dam components, which gave estimates of about 0·50 and 0·20, and the within-dam components, which gave estimates of about −0·10 and −0·03, for birth and weaning weights respectively. The regression of calf weaning weight on dam weight at the beginning of the breeding season averaged 0·12 kg per kg, indicating that dams 50 kg lighter than average weaned calves 6 kg lighter than average. The regressions of calf weaning weight on dam weaning weight were about half this value and the correlations of these two weights were also reduced, particularly in the Angoni breed which had a small between-dam component. These reductions occurred because all breeds had negative correlations between calf weaning weight and dam live-weight change measured over the major part of the lactation period. These negative correlations suggest that the heaviest weaners were produced by the dams losing most weight during lactation. Both the correlation and the regression of calf weaning weight on dam live-weight change were lower in the Barotse breed than in the Angoni and Boran breeds, and the correlation in Barotses was not significantly different from zero.

During the study most dams had live weights recorded in more than one season and gave birth to and reared more than one calf. The repeatabilities of these dam traits were estimated as the intra-class correlation of weight records of the same dam, V_b/V_b+V_w, where V_b is the between-dam component of variance and V_w the within-dam component of variance. The estimates are presented in Table 6 and averaged 0·29 and 0·52 for calf birth and weaning weights respectively, and were highest in the Angoni breed. All dam live-weight repeatabilities exceeded 0·80.

TABLE 6

Repeatabilities of calf and dam live weights

Dam breed	BW	WW	BBS	DWW
Angoni	0·326	0·625	0·875	0·821
Barotse	0·246	0·467	0·856	0·855
Boran	0·289	0·460	0·814	0·837
Av. s.e.	0·060	0·050	0·016	0·021

DISCUSSION

There has been little published work from Zambia on the reproductive performance of beef cattle. Rakha, Hale and Igboeli (1970) have reported the age of puberty of Angoni, Africander, Mashona and Hereford heifers, and Rakha, Igboeli and King (1971) have compared the calving intervals, gestation and *post partum* periods of the same breeds. In the latter comparison, which had similar management conditions to the present study, the Angoni had the shortest mean calving interval and highest calving rate, 83%. The apparent excellent fertility of the Angoni breed was confirmed in this present comparison in which the two indigenous breeds, Angoni and Barotse, and the Boran breed introduced from Kenya, all exceeded a calving rate of 75%. This calving rate is well above the national average for commercial beef herds.

The major influence on calving rate for each of the breeds was the difference between years. In the extreme case, calving percentages were 23·0, 28·2 and 14·4 lower in 1970 than in 1966 for the Angoni, Barotse and Boran breeds respectively. Some of these year differences will have been due to the effect of rainfall variation on the quantity and quality of available herbage. A confounding effect resulted from the breeding herds being unable to graze the Kafue peneplain during the 1969 dry season because of its prolonged flooding. An additional factor contributing to the depressed calving percentages in 1970 may have been the increased stocking rate resulting from the expansion of the experimental programme. An examination of individual sire records did not indicate any important decline in sire fertility during the programme.

The effect of dam status at mating on calving percentage varied with breed. Under these favourable conditions it was found that most lactating Angoni and Boran dams were able to conceive in successive years. This contrasted with the performance of dams of the sanga breed, Barotse, which were more likely to calve if dry at mating, suggesting that lactation had a greater effect on the fertility of this breed. Dams lactating at mating had lower calving rates than dry dams in the comparison of Brahman and Santa Gertrudis reported by Willis and Wilson (1974), and in the comparison of

Africander, Tswana, Tuli and Brahman dams in Botswana (Buck, Light, Rutherford, Miller, Rennie, Pratchett, Capper and Trail, 1976).

Among the lactating dams of $\geqslant 5$ years in the Angoni and Barotse, but not the Boran breed, the earlier a dam had calved during the previous season the more likely she was to reconceive. In the Botswana study, Buck *et al.* (1976) found the same relationship in a larger body of data although breed differences were not examined. Some differential breed effects for dam status on calving percentage may have occurred in the Uganda comparison of Ankole, Boran and Zebu dams reported by Trail, Sacker and Fisher (1971), who found significant pooled interactions for calving percentage. The breed differences observed between the Angoni, Barotse and Boran may be the result of variations in the period required for complete involution of the uterus and the resumption of ovarian activity. In their study Rakha *et al.* (1971) report a range of breed estimates for this *post partum* period, including one of $79{\cdot}9 \pm 19{\cdot}2$ days for Angoni dams. If major breed differences are confirmed it may be warranted to delay the start of the breeding season for breeds with above average *post partum* periods. Some increases in sire : dam ratios may be required if the date for the completion of the breeding season is to remain unchanged.

In the Botswana data fertility declined in dams $\geqslant 8$ years (Buck *et al.*, 1976), whereas in the present study fertility did not decline until dams reached 13 years, a finding consistent with that of Willis and Wilson (1974). Age did not have a significant effect on calving percentage in the Ugandan comparison which included Boran dams (Trail *et al.*, 1971), and the low calving rate of the Zambian Boran heifers is consistent with the dam status results in that study. Angoni calving rates at 3 and 4 years were comparable with those of the older Angoni dams, indicating that with good management mating at 2 years is practicable.

Sire breed effects on calving rate under African conditions do not appear to have been reported previously. While their effects were not generally significant, there was some indication that higher calving rates were achieved when Angoni and Barotse dams were mated to sires of their own breed. Unfortunately, there were too few records to give valid estimates of the effect of *Bos taurus* sire breeds which, with the exception of the Hereford matings with Barotse dams and Friesian matings with Angoni dams, gave lower calving rates than sires of the African breeds. In the southern United States several experiments have shown that sires of *Bos taurus* breeds give significantly lower calving percentages than Brahman sires (Turner, Farthing and Robertson, 1968; Peacock, Koger, Kirk, Hodges and Warnick, 1971; Peacock, Koger, Crockett and Warnick, 1977). In Zambia *Bos taurus*-sired crossbreds have been shown to have advantages for live and carcass weights (Thorpe, Cruickshank and Thompson, 1979). Any depression in calving rate caused by *Bos taurus* sires, or managment costs incurred to offset them, would reduce their superiority as sires of F_1 slaughter progeny. Accurate estimates of these sire breed effects on calving rates and estimates of the reproductive performance of *Bos taurus*-sired F_1 female progeny are required.

Pre-weaning mortality rates were generally lower than those reviewed by Preston and Willis (1974), and were of similar magnitude to Botswanan estimates for mortality to 1 year (Trail, Buck, Light, Rennie, Rutherford, Miller, Pratchett and Capper, 1977). Unpublished estimates from an earlier breed comparison in Zambia (Rakha *et al.*, 1971) gave rates of 3·8, 4·1 and

8·9% for purebred progeny of Angoni, Hereford and Africander dams respectively. In the present study, the lower Angoni pre-weaning mortality rate increased the superiority shown by the breed for calving rate, such that Angoni dams had a marked advantage in net calf crop over the introduced zebu breed Boran, with the sanga breed, Barotse, intermediate. The superiority of the small Angoni breed over the Barotse was maintained under less favourable conditions in a later comparison in which the Boran net calf crop exceeded that of Barotse dams (W. Thorpe, D. K. R. Cruickshank and R. Thompson, unpublished work).

The lower progeny birth and weaning weights of dams calving in successive years and the higher weaning weights of intermediate age Barotse dams were consistent with the results of many reports, including that of Sacker, Trail and Fisher (1971). The absence of other effects on calf weights is further indication of the favourable conditions under which the study was conducted. The sex differences were similarly consistent with the findings of Sacker *et al.* (1971), with male calves heavier at birth than females by about 1·7 kg or 7%, and at weaning by about 14 kg or 8%. Additional variation in calf weaning weight in all breeds resulted from the effect of date of birth. Period of birth effects for calf weaning weight were only important in the Angoni breed, although related trends were shown in the other breeds. These effects indicated that management practices aimed at increasing the proportion of early calvers will not only increase the likelihood of the dam's reconception but may also have advantages for calf weaning weight beyond that established from date of birth regressions.

Correlations and regressions for calf and dam live weights and live-weight changes were very similar to those reported by Trail *et al.* (1971). It appears that while dam and calf weights are positively correlated, the heaviest weaners were those reared by dams which lost most weight during the suckling period. The repeatabilities (Table 6) indicate upper limits of heritabilities and are in line with the values reviewed by Preston and Willis (1974).

While dam live weights and live-weight changes were related to progeny live weights, there were no significant relationships between dam live weight and live-weight change and subsequent calving rate. The validity of the concept proposed by Lamond (1970), that there is a target body weight below which the ability of a cow to conceive is decreased, has been substantiated in Zimbabwe-Rhodesia by Elliot (1964), Ward (1968), Richardson, Oliver and Clarke (1975) and Steenkamp, van der Horst and Andrew (1975), and in Botswana by Buck *et al.* (1976). The absence of an effect of live weight on calving rate in the present study suggests that nutrition was sufficient to maintain dams at live weight above the target weight. Lactating dams had significantly lighter weights than dry dams at the beginning of the breeding season and Topps (1977) has noted that in most tropical environments the higher fertility of dry cows is associated with and confounded by such differences. The finding in the Angoni and Boran breeds that lactating dams tended to be more fertile than dry dams suggests that lactation is not an important influence on fertility in these breeds when live weights are adequate. On the other hand, lactation tended to depress the fertility of the Barotse dams. It is possible that some hormonal action associated with lactation influenced fertility in the sanga breed but not in the two zebu breeds.

Year as a source of variation influencing calving rate was confounded in 1970 with the unavailability of flats grazing during the preceding dry season.

The lack of flats grazing will have contributed to the live-weight loss between the 1969 autumn peak and the following mid breeding season weight, which was greater than in any other year. However, the weight loss was still relatively small, and the estimates of the relationship between live-weight change and subsequent calving rate indicate that this weight loss is unlikely to account for a major part of the drop in calving rate between 1969 and 1970. Alternatively, the availability of fresh herbage on the flats grazing during the dry season can be equated to protein supplementation, which has been shown to have a beneficial effect on dam fertility independent of live-weight change in sheep (Lindsay, 1976) and in cattle (Ward, 1968; Capper, Pratchett, Rennie, Light, Rutherford, Miller, Buck and Trail, 1977). It seems probable that the lack of the 'protein supplementation' provided by flats grazing contributed to the depressed calving percentage in 1970. As flats grazing is not generally available to ranch herds in Zambia some quantification of the calving rate response to dry season protein supplementation seems necessary, particularly as an alternative management technique, early weaning, has had no beneficial effect on subsequent calving rates in Zambia (Rakha *et al.*, 1971) and in Zimbabwe-Rhodesia (Richardson *et al.*, 1975). Legume-reinforced veld (van Rensburg, 1969) may prove a more economic source of dry season protein than hand feeding.

ACKNOWLEDGEMENTS

The authors thank the Director of Agriculture for the provision of experimental facilities and acknowledge the technical assistance of C. D. Slater, P. J. Hatakwati and all Animal Husbandry Research staff at Mazabuka, and the help given by the Veterinary Department. Thanks are also due to Professor J. H. D. Prescott, Edinburgh School of Agriculture, Professor D. J. Finney, ARC Unit of Statistics, Dr G. Wiener and Mr I. L. Mason. The UK Ministry of Overseas Development financed the reporting of the results through research scheme R3390.

REFERENCES

Buck, N. G., Light, D., Rutherford, A., Miller, M., Rennie, T. W., Pratchett, D., Capper, B. S. and Trail, J. C. M. 1976. Environmental factors affecting beef cow reproductive performance in Botswana. *Anim. Prod.* **23:** 357–363.

Capper, B. S., Pratchett, D., Rennie, T. W., Light, D., Rutherford, A., Miller, M., Buck, N. G. and Trail, J. C. M. 1977. Effects of rumen stimulatory licks on reproductive performance and live-weight gain of beef cattle in Botswana. *Anim. Prod.* **24:** 49–55.

Elliot, R. C. 1964. Some nutritional factors affecting the productivity of beef cattle in Southern Rhodesia. *Ph.D. Thesis, Univ. London.*

Lamond, D. R. 1970. The influence of undernutrition on reproduction in the cow. *Anim. Breed. Abstr.* **38:** 359–372.

Lindsay, D. R. 1976. The usefulness to the animal producer of research findings in nutrition on reproduction. *Proc. Aust. Soc. Anim. Prod.* **11:** 217–224.

Mason, I. L. and Maule, J. P. 1960. *The Indigenous Livestock of Eastern and Southern Africa.* Commonwealth Agricultural Bureaux, Farnham Royal, Buckinghamshire.

Peacock, F. M., Koger, M., Crockett, J. R. and Warnick, A. C. 1977. Reproductive performance and crossbreeding Angus, Brahman and Charolais cattle. *J. Anim. Sci.* **44:** 729–733.

Peacock, F. M., Koger, M., Kirk, W. G., Hodges, E. M. and Warnick, A. C. 1971. Reproduction in Brahman, Shorthorn and crossbred cows on different pasture programs. *J. Anim. Sci.* **33:** 458–465.

Preston, T. R. and Willis, M. B. 1974. *Intensive Beef Production.* 2nd ed. Pergamon, Oxford.

Rakha, A. M., Hale, D. and Igboeli, G. 1970. The age of puberty in local breeds of cattle in central Africa. *J. Reprod. Fert.* **22:** 369–370.

RAKHA, A. M., IGBOELI, G. and KING, J. L. 1971. Calving interval, gestation and post-partum periods of indigenous central African cattle under a restricted system of breeding. *J. Anim. Sci.* **32:** 507–509.

RICHARDSON, F. D., OLIVER, J. and CLARKE, G. P. Y. 1975. Analyses of some factors which affect the productivity of beef cows and of their calves in a marginal rainfall area of Rhodesia. 1. Factors affecting calving rate. *Anim. Prod.* **21:** 41–49.

SACKER, G. D., TRAIL, J. C. M. and FISHER, I. L. 1971. Crossbreeding beef cattle in western Uganda. 2. Environmental influences on body weight. *Anim. Prod.* **13:** 143–152.

STEENKAMP, J. D. G., VAN DER HORST, CONNY and ANDREW, M. J. A. 1975. Reconception in grade and pedigree Africander cows of different sizes: postpartum factors influencing reconception. *S. Afr. J. Anim. Sci.* **5:** 103–110.

THORPE, W., CRUICKSHANK, D. K. R. and THOMPSON, R. 1979. The growth and carcass characteristics of crosses of Hereford and Friesian with Angoni, Barotse and Boran cattle in Zambia. *J. agric. Sci., Camb.* **93:** 423–430.

TOPPS, J. H. 1977. The relationship between reproduction and undernutrition in beef cattle. *Wld Rev. Anim. Prod.* **13** (2): 43–49.

TRAIL, J. C. M., BUCK, N. G., LIGHT, D., RENNIE, T. W., RUTHERFORD, A., MILLER, M., PRATCHETT, D. and CAPPER, B. S. 1977. Productivity of Africander, Tswana, Tuli and crossbred beef cattle in Botswana. *Anim. Prod.* **24:** 57–62.

TRAIL, J. C. M., SACKER, G. D. and FISHER, I. L. 1971. Crossbreeding beef cattle in western Uganda. 1. Performance of Ankole, Boran and Zebu cows. *Anim. Prod.* **13:** 127–141.

TURNER, J. W., FARTHING, B. R. and ROBERTSON, G. L. 1968. Heterosis in reproductive performance of beef cows. *J. Anim. Sci.* **27:** 336–338.

VAN RENSBURG, H. J. 1969. Legume/grass pastures in Zambia. *Fmg Zambia* **4:** 2.

WARD, H. K. 1968. Supplementation of beef cows grazing on veld. *Rhod. J. agric. Res.* **6:** 93–101.

WILLIS, M. B. and WILSON, A. 1974. Comparative reproductive performance of Brahman and Santa Gertrudis cattle in a hot humid environment. 1. Fertility and descriptive statistics. *Anim. Prod.* **18:** 35–42.

(*Received* 13 *July* 1979)

33

Genetic and environmental influences on beef cattle production in Zambia. 2. Sire weights for age of purebred and reciprocally crossbred progeny

by

W. Thorpe, D.K.R. Cruickshank and R. Thompson

Anim. Prod. 1980, **30**: 235–243

GENETIC AND ENVIRONMENTAL INFLUENCES ON BEEF CATTLE PRODUCTION IN ZAMBIA

2. LIVE WEIGHTS FOR AGE OF PUREBRED AND RECIPROCALLY CROSSBRED PROGENY

W. THORPE AND D. K. R. CRUICKSHANK

Central Research Station, Mazabuka, Zambia

AND

R. THOMPSON

ARC Unit of Statistics, University of Edinburgh, Edinburgh EH9 3JZ

ABSTRACT

Live weights from birth to 3·5 years are reported for beef cattle reared under ranching conditions in Zambia. The 809 cattle were purebred Africanders, Angonis, Barotses and Borans and the reciprocal crossbreds of the latter three breeds born in 2 years. All animals born in the 1st year and half the males born in the 2nd year grazed natural grassland. The remaining males and all females born in the 2nd year received, in addition, dry season supplementary feed from 1·5 years of age.

The interaction of genotype with year-of-birth was important but not the interactions of genotype with management or sex.

Purebred progeny of the introduced Africander breed were heavier than the progeny of the indigenous Angoni and Barotse breeds in both year-of-birth groups, but only heavier than progeny of the introduced Boran breed in the first group. On average, the Africander progeny had live-weight advantages of about 16% and 10%, and the Boran progeny advantages of about 12·5% and 5·5% over the purebred Angoni and Barotse progeny respectively. Heterosis estimates tended to increase with age, reaching levels of about 5 to 6% in the Barotse/Boran crosses at and after 1·5 years. Heterosis was not shown by the other crosses. The Barotse and Boran breeds had similar maternal effects which were superior to those of the Angoni breed.

INTRODUCTION

A breed evaluation programme with *Bos indicus*, *Bos taurus* and sanga cattle under ranch conditions in Zambia comprised each of the three phases of the beef cattle production cycle. Breed descriptions and the reproductive performance of purebred dams of the Angoni, Barotse and Boran breeds have been given by Thorpe, Cruickshank and Thompson (1980a). The present paper reports live weights from birth to 3·5 years of age for the purebred and

reciprocally crossbred progeny of these dams and for the purebred progeny of Africander dams. The final phase of the production cycle, the carcass characteristics of the male castrate progeny considered in this paper, is reported by Thorpe, Cruickshank and Thompson (1980b).

MATERIAL AND METHODS

The background to the programme has been described by Thorpe *et al.* (1980a). Progeny for the comparison were born in 1967 and 1969 between late September and mid-December, and the latter part of the calving season overlapped with the beginning of the rainy season. Male calves were castrated in April and all calves were weaned at about 8 months of age over a 3-day period in early June. Progeny were weaned into pens and fed there for about 4 weeks before returning to extensive *Hyparrhenia* veld grazing. No supplementary feeds except rock salt were available prior to weaning. During the summer rainfall period, November to March, ample grazing gave rapid live-weight gains. In the dry season, April to October, the mature herbage provided maintenance or submaintenance nutrition.

All progeny were bred and reared at Mazabuka apart from the purebred Africanders. In 1967 Africander dams of unknown age and previous parous state were borrowed with calves at foot from a neighbouring commercial ranching company. The calves had a mean age of about 2 months and were contemporary with the Mazabuka-born progeny but had no birth data. In the 2nd year Africander weaners, contemporary with the group born in 1969, were purchased from the same company. The Mazabuka-bred progeny were sired by 9 Angoni, 8 Barotse and 9 Boran males (bulls) in single sire herds in each of which the three breeds were represented. In this way all nine possible breed combinations were produced and sire and heterosis effects were not confounded. Progeny with complete records totalled 809 and included 158 Africanders. The number of records for progeny of the other breeds and crosses ranged from 65 to 81.

Up to 1·5 years of age both year-of-birth groups had the same management and all contemporaries grazed natural veld in a single herd. Thereafter the 1967-born progeny continued to graze veld without dry season supplementary feed and live weights up to 3·5 years of age were recorded. Up to 2·5 years of age about half of the 1969-born castrated males (steers) grazed veld with no supplementary feed. By contrast, the remaining steers and all young adult females (heifers) were fed supplements for 145 days after 1·5 years of age, that is during their second dry season. The feeding period immediately preceded the start of the rains, which gave fresh growth of herbage and hence a rising plane of nutrition. Immediately after the rains the previously fed steers were again fed supplements for a further 40 days prior to slaughter at 2·5 years of age. The heifers and previously unsupplemented steers were restricted to veld grazing.

The records were analysed as live weights-for-age at birth and weaning, and at 6-monthly intervals from 0·5 to 3·5 years for the 1967-born and from 0·5 to 2·5 years for the 1969-born progeny. The post-weaning live weights were recorded on the day approximating to the date on which the mean age of the group corresponded to 1·0, 1·5, 2·0 years, etc. In the absence of supplementary feeding the live weights at 1·0, 2·0 and 3·0 years generally corresponded to the minimum live weight recorded at the end of the dry

season. The half-yearly weights, generally recorded in late May, were usually the maximum live weights attained after the rainy season. Animals were not starved before weighing.

Preliminary least squares analyses fitted a model which included the interaction of sire breed, dam breed and sex. There was no important differential heterosis in the two sexes and the combined population was re-analysed omitting the sex interaction effects. Further analyses indicated that year-of-birth × genotype interaction effects were significant for the live weights from birth to 1·5 years of age. Consequently, birth to 1·5 year live weights are reported separately for each year-of-birth. The model used in the analyses fitted the effects period of birth (<294, 294-314, $\geqslant 315$ days), dam age (3, 4, 5 and 6, 7 and 8, 9, $\geqslant 10$ years), dam previous parous state (heifer, calved previous year, barren previous year), sex (female, male), genotype (Africander, Angoni, Barotse, Boran, Angoni × Barotse, Barotse × Angoni, Angoni × Boran, Boran × Angoni, Barotse × Boran, Boran × Barotse) and a regression for date of birth. Later live weights for the 1967-born group were analysed using the same model. The model for the analyses of the 2·0 and 2·5 year of age live weights for the group born in 1969 also included the effects of management and genotype × management interaction. Year-of-birth was the only pre-weaning information available for the Africander progeny. The Africander results therefore included variation due to date and period of birth, dam age and previous parous state, and any variation resulting from the different pre-weaning environments. Heterosis was estimated as the difference between the mean of the reciprocal crossbreds and the midparent mean.

RESULTS

The environmental effects associated with the individual's date of birth and sex gave significant differences at most ages. Other environmental effects were only occasionally significant. In both years male progeny were 2, 15 and 20 kg heavier than females at birth, weaning and 1·5 year respectively, and the sex difference at all ages was about 8%. In neither year did date of birth affect birth weight, but calves born 30 days later than average were lighter by 23 and 16 kg at weaning, 18 and 15 kg at 1·0 year, and 20 and 18 kg at 1·5 years for the 1967- and 1969-born groups respectively. The weight advantage of the earlier-born progeny reduced with age, and tended to be greater at the end of the rains than at the end of the dry season.

Genotypic variation was significant at all ages and in both years. The least squares means and heterosis estimates for the live weights from birth to weaning and from 1·0 to 3·5 years of age are presented in Tables 1 and 2 respectively.

Purebred comparisons. In both years and at all ages purebred Angoni progeny were significantly lighter than the purebred Barotse and Boran progeny. The live weights of the latter breeds were not significantly different in the 1967-born group except at 2·0 and 2·5 years of age, although the Boran live weights were higher at all ages, a superiority which was generally highly significant in the 1969-born group.

At 6 months and at weaning (Table 1) the Africander breed was significantly heavier than the Angoni and Barotse, but was not significantly heavier than the Boran. At the later ages (Table 2) 1967-born Africander progeny were significantly heavier than the other three purebreds. The 1969-born

TABLE 1

Genotype means and heterosis estimates for live weights from birth to weaning

	Live weights (kg)					
	Birth		6 months		Weaning	
	1967	1969	1967	1969	1967	1969
Purebreds						
Africander	—	—	162·5a	—	173·8a	—
Angoni	21·2	22·3	135·6	129·7	146·5	142·2
Barotse	24·8a	24·2	155·7a	138·4	171·3a	154·7
Boran	25·0a	25·8	158·8a	152·2	174·1a	169·2
Approx. s.e. of difference between purebred means	0·73	0·71	3·75	3·50	4·31	3·92
Reciprocal crossbreds						
Angoni × Barotse	24·[illegible]	[illegible]	151·2	138·4	166·9	154·7
Barotse × Angoni	21·9	[illegible]	[illegible]·7	128·0	149·8	141·1
Heterosis						
units	0·2	[illegible]	−1·2	−0·9	−0·6	−0·6
(%)	0·9	0·5	−0·8	−0·7	−0·4	−0·4
Angoni × Boran	24·3a	23·6a	152·3	140·9a	166·4	156·9a
Boran × Angoni	23·0a	24·1a	136·7	137·4a	147·9	151·7a
Heterosis						
units	0·6	−0·3	−2·8	−1·8	−3·1	−1·4
(%)	2·6	−1·1	−1·9	−1·3	−2·0	−0·9
Barotse × Boran	23·4	24·8	156·6a	149·1a	174·2a	167·4a
Boran × Barotse	27·0	28·3	162·5a	152·8a	179·0a	170·7a
Heterosis						
units	0·3	1·5*	3·9	5·6	3·9	7·1
(%)	1·1	6·1	2·3	3·9	2·3	4·4
Approx. s.e. of difference between reciprocal crossbred means	0·60	0·73	3·92	3·62	4·50	4·05
Approx. s.e. of heterosis estimates	0·75	0·73	3·88	3·62	4·46	4·05

* Significantly different from zero ($P < 0·05$).

a Those means within purebred and reciprocal crossbred classes followed by the same letter do not differ significantly ($P > 0·05$).

Africanders were significantly heavier post-weaning than the Angonis and Barotses but had live weights similar to the Borans.

Heterosis estimates. Heterosis estimates at all live weights from birth to 3·5 years of age were non-significant for the Angoni/Barotse and Angoni/Boran crossbreds. The Angoni/Barotse reciprocals had zero heterosis at early ages, increasing to levels near significance at 3·0 and 3·5 years of age. The heterosis estimates for the Angoni/Boran reciprocals showed a similar though less marked increase with age. Significant heterosis was shown by the Barotse/Boran reciprocals for birth weight in 1969 but not in 1967. Later ages in both groups had positive, non-significant heterosis which increased to reach significance at 1·5 years of age and at all subsequent ages (Table 2).

Reciprocal crossbred comparisons. Live weight-for-age estimates of reciprocal crossbred differences give comparisons of maternal effects. For example, in both years crossbred progeny from Barotse dams compared with those from Angoni dams had significantly higher weights to 1·0 year of age,

TABLE 2

Genotype means and heterosis estimates for post-weaning live weight

Years of age	1·0		1·5		2·0		2·5		3·0	3·5
	1967	1969	1967	1969	1967	1969	1967	1969	1967	1969
Purebreds										
Africander	162·7	166·8a	256·1	232·7a	237·8	232·3b	338·2	339·4a	286·8	411·6
Angoni	129·9	138·6	208·7	202·5	192·7	213·8a	274·5	295·0	238·3	328·1
Barotse	149·2a	151·8	235·0a	216·0	215·3a	220·5ab	307·5	313·8	255·3	367·5a
Boran	150·0a	160·0a	241·1a	234·3a	223·8a	252·0	318·9	338·3a	270·7	375·4a
Approximate s.e. of difference between purebred means	3·45	3·63	4·94	4·56	4·38	6·36	5·94	7·09	5·49	7·19
Reciprocal crossbreds										
Angoni × Barotse	145·2	153·0	229·7a	222·9	212·2a	230·1	297·5a	313·7	258·7a	358·7a
Barotse × Angoni	137·4	137·8	224·3a	203·1	207·7a	21[illegible]	298·7a	299·2	255·2a	363·3a
Heterosis										
units	1·8	0·2	5·2	3·7	5·9	[illegible]	7·1	2·1	10·2	13·2
(%)	1·3	0·2	2·3	1·8	2·9	[illegible]	2·4	0·7	4·1	3·8
Angoni × Boran	145·2	151·6a	234·7	222·0a	215·9	23[illegible]2a	307·1	321·1a	264·1	364·3a
Boran × Angoni	127·5	145·5a	212·5	220·0a	200·9	235·7a	291·1	328·2a	250·6	349·1a
Heterosis										
units	−3·7	−0·8	−1·3	2·6	0·1	4·6	2·4	8·0	2·9	5·0
(%)	−2·6	−0·5	−0·6	1·2	0·0	2·0	0·8	2·5	1·1	1·4
Barotse × Boran	155·4a	159·2a	252·3a	234·0	230·2a	245·0a	328·4a	344·1	279·3a	387·2a
Boran × Barotse	155·5a	166·3a	166·3a	244·9	228·7a	257·0a	327·8a	359·3	277·5a	390·9a
Heterosis										
units	5·8	6·9	12·7*	14·3**	9·9*	14·7*	14·9*	25·6***	15·4**	17·6*
(%)	3·9	4·4	5·3	6·4	4·5	6·2	4·8	7·9	5·9	4·7
Approximate s.e. of difference between reciprocal crossbred means	3·60	3·75	5·16	4·72	4·58	6·63	6·20	7·38	5·74	7·51
Approximate s.e. of heterosis estimates	3·57	3·75	5·12	4·72	4·54	6·52	6·15	7·27	5·69	7·45

a, b. Those means within purebred and reciprocal crossbred classes followed by the same letter do not differ significantly ($P > 0·05$).

clearly demonstrating the superiority of the Barotse maternal effect. The advantage persisted to the later ages in the 1969 group. There were no significant differences between the reciprocal crossbreds at 1·5 years of age or later in the 1967-born group.

The reciprocal crossbred calves born to the Boran and Angoni dams in both years did not differ significantly for birth weight. In the 1967 group, progeny born to Boran dams had higher live weights at 6 months and weaning, and the significant advantage was maintained up to 3·0 years of age. The superior maternal influence of the Boran was again shown in the 1969 group but the differences were not statistically significant.

In both years crossbred calves born to Barotse dams were significantly heavier at birth than the calves born to Boran dams (Table 1). Thereafter there were generally no significant differences between the reciprocal crossbred live weights, suggesting that the Barotse dams provided a superior uterine environment, but that the breeds produced similar quantities of milk.

DISCUSSION

The predominant indigenous breeds of southern Africa will inevitably form the basis of beef production improvement programmes (Maule, 1973). Results in Botswana (Trail, Buck, Light, Rennie, Rutherford, Miller, Pratchett and Capper, 1977) and in Zimbabwe-Rhodesia (Department of Research and Specialist Services, 1978) have shown variation in productive performance between a wide selection of breeds and crosses under ranching conditions which, together with the present results, indicate the potential that exists for breed replacement and crossbreeding.

The value of crossbreeding depends on both the additive and non-additive gene effects on performance. The present results indicate that, even if only additive effects are assumed, Africander sires would give live weight advantages of about 8% and 5% respectively for crossbred progeny from Angoni and Barotse dams. Less encouraging are the findings that the Africander breed has a lower fertility than most other breeds available in southern Africa (Rakha, Igboeli and King, 1971; Trail *et al.*, 1977; W. Thorpe, D. K. R. Cruickshank and R. Thompson, unpublished work). If reliable estimates indicate that the Africander gives advantageous heterosis for maternal and fitness characters then it may be a valuable crossing sire for Zambia. Maule (1973) has noted the absence of these estimates.

Compared with the Zambian breeds the fertility of the east African Boran is more competitive than the Africander (W. Thorpe, D. K. R. Cruickshank and R. Thompson, unpublished work). In addition the Boran has advantages for live weight of about 12·5% and 5·5% respectively over the purebred Angoni and Barotse progeny at 2·5 and 3·0 years of age, and these advantages are sustained to carcass weight (Thorpe *et al.*, 1980b). Despite the lack of heterosis between the Boran and Angoni, which are both short-horned zebu breeds (Mason and Maule, 1960), the additive effects on live weight show the Boran to be a useful crossing sire breed with the Angoni. There was no heterosis for the Angoni/Barotse crosses despite their distinct zebu and sanga breed types. The larger dams, the Barotses, had significantly superior maternal influence, probably due to their larger size, as did Boran dams when compared with the Angoni.

More encouraging was the significant heterosis shown by Barotse/Boran

crosses at post-weaning ages. The two breeds had similar maternal influences on progeny live weights and the reciprocals were significantly heavier by about 10% than the lighter Barotse parent breed, and heavier by about 6% than the superior Boran parent breed. The heterosis level, between 5 and 6%, was within the range of estimates reviewed by Preston and Willis (1974), but lower than expected according to Plasse (1976) for *Bos taurus* × *Bos indicus* cattle. The crossbred advantage was smaller than that given by Brahman bulls, and larger than by Africander bulls crossed with Tswana dams in Botswana (Trail *et al.*, 1977). Mason and Maule (1960) state that the Tswana and Barotse breeds are closely related. Crossing these breeds with *Bos taurus* bulls has given larger live weight advantages in Botswana (Trail *et al.*, 1977), and in Zambia (Thorpe, Cruickshank and Thompson, 1979), than that shown here in crosses with the Boran.

The lack of heterosis in the Angoni/Barotse crosses was more surprising than its absence in the Angoni/Boran crosses, and illustrates the danger of applying generalized estimates in the design of breeding programmes. American evidence suggests that heterosis may be higher at better management levels (Lasley, Sibbit, Langford, Comfort, Dyer, Krause and Hedrick, 1973), and it is possible that the low production levels reported here (live-weight gains of less than 0·4 kg/day) may have suppressed heterosis effects. Long and Gregory (1975) reported larger heterosis effects on live-weight gain at higher rates of gain, i.e. at better management levels, or at younger ages. It was not possible to separate the factors in their data. For the breeds examined here there is a suggestion that the non-additive effects increased with increasing age. For example, significant heterosis between the Barotse and Boran breeds was only found at the post-weaning ages. Nitter (1978) has found in sheep that a decreasing maternal influence on growth due to age seems to be reflected in an increase in individual heterosis, and has suggested that the purebred maternal environment may be a limiting factor on the full expression of the growth potential of F_1 progeny at early ages. By contrast, Sacker, Trail and Fisher (1971b) report significant heterosis in birth weight and at weights up to 12 months of age for crosses of Boran and Red Poll in Uganda. An 'environmental insufficiency' (Nitter, 1978) inhibiting the expression of Red Poll dam performance is likely to have contributed to these estimates, and the influence of non-additive gene effects is uncertain. A further source of variation, the possibility that heifers have higher heterosis levels than steers, was not examined in the Ugandan data. Preston and Willis (1974) quote early American reports showing this phenomenon, but it was not found in the data of Long and Gregory (1975) or in the present analyses.

The interaction of genotype and management was not significant in the 1969-born group. The small subclass numbers made it unlikely that any interaction of practical importance would be detected. However, the larger year-of-birth groups did give significant interactions with genotype which indicated that year-of-birth was an important influence on the magnitude of the breed and crossbred differences for live weight. Interactions have also been reported by Trail (1969) and Thorpe *et al.* (1979). Variation due to the different sires and dams sampled in the two breeding seasons and the interaction of genotypes and year environments will have contributed to these effects. The interactions indicate the necessity for breed comparisons to be conducted over a number of years when the experimental environment has large annual and monthly variation in rainfall.

As would be expected from the results of Thorpe *et al.* (1980a), the environmental sources of variation, dam age and previous parous state were generally unimportant influences on progeny live weights. Sacker, Trail and Fisher (1971a), in a Ugandan comparison which included the Boran, found only occasionally significant effects of dam age and previous parous state, but the seasonal effect, period of birth, had highly significant effects on virtually all body weights. In the present experiment, where the date of birth covariate accounted for the significant variation due to differences in age at the time of weighing, period effects were unimportant. The sex differences reported here were of a similar magnitude to those reported by Sacker *et al.* (1971a) and many others in the literature.

It is clear from the comparison that both imported Africander and Boran breeds have superior live weights to those of the Zambian breeds, Angoni and Barotse. Moreover, crossbreeding the Zambian breeds with the Boran has live-weight advantages for both breeds, and particularly for the Barotse because of heterosis effects. Together with the competitive reproductive performance of Boran crosses (W. Thorpe, D. K. R. Cruickshank and R. Thompson, unpublished work), these results suggest that the Boran could have a significant role to play in the improvement of Zambia's indigenous cattle populations.

REFERENCES

Department of Research and Specialist Services. 1978. *A.Rep. 1977, Div. Livestock and Pastures.* Rhodesian Government Printer, Salisbury.

Lasley, J. F., Sibbit, B., Langford, L., Comfort, J. E., Dyer, A. J., Krause, G. F. and Hedrick, H. B. 1973. Growth traits in straightbred and reciprocally crossed Angus, Hereford and Charolais steers. *J. Anim. Sci.* **36:** 1044–1056.

Long, C. R. and Gregory, K. E. 1975. Heterosis and management effects in postweaning growth of Angus, Hereford and reciprocal cross cattle. *J. Anim. Sci.* **41:** 1563–1571.

Mason, I. L. and Maule, J. P. 1960. *The Indigenous Livestock of Eastern and Southern Africa.* Commonwealth Agricultural Bureaux, Farnham Royal, Buckinghamshire.

Maule, J. P. 1973. The role of the indigenous breeds for beef production in southern Africa. *S. Afr. J. Anim. Sci.* **3:** 111–130.

Nitter, G. 1978. Breed utilization for meat production in sheep. *Anim. Breed. Abstr.* **46:** 131–143.

Plasse, D. 1976. The possibility of genetic improvement of beef cattle in developing countries with particular reference to Latin America. In *Beef Cattle Production in Developing Countries* (ed. A. J. Smith), pp. 308–331. Edinburgh University Press, Edinburgh.

Preston, T. R. and Willis, M. B. 1974. *Intensive Beef Production.* 2nd ed. Pergamon, Oxford.

Rakha, A. M., Igboeli, G. and King, J. L. 1971. Calving interval, gestation and post-partum periods of indigenous central African cattle under a restricted system of breeding. *J. Anim. Sci.* **32:** 507–509.

Sacker, G. D., Trail, J. C. M. and Fisher, I. L. 1971a. Crossbreeding beef cattle in western Uganda. 2. Environmental influences on body weight. *Anim. Prod.* **13:** 143–152.

Sacker, G. D., Trail, J. C. M. and Fisher, I. L. 1971b. Crossbreeding beef cattle in western Uganda. 6. A note on hybrid vigour in Red Poll-Boran crosses. *Anim. Prod.* **13:** 181–184.

Thorpe, W., Cruickshank, D. K. R. and Thompson, R. 1979. The growth and carcass characteristics of crosses of Hereford and Friesian with Angoni, Barotse and Boran cattle in Zambia. *J. agric. Sci., Camb.* **93:** 423–430.

Thorpe, W., Cruickshank, D. K. R. and Thompson, R. 1980a. Genetic and environmental influences on beef cattle production in Zambia. 1. Factors affecting weaner production from Angoni, Barotse and Boran dams. *Anim. Prod.* **30:** 217–234.

THORPE, W., CRUICKSHANK, D. K. R. and THOMPSON, R. 1980b. Genetic and environmental influences on beef cattle production in Zambia. 3. Carcass characteristics of purebred and reciprocally crossbred castrated males. *Anim. Prod.* **30:** 245–252.

TRAIL, J. C. M. 1969. A study of performance traits of crossbred cattle in a new beef cattle ranching area in Uganda. *Ph.D. Thesis, Univ. Edinburgh.*

TRAIL, J. C. M., BUCK, N. G., LIGHT, D., RENNIE, T. W., RUTHERFORD, A., MILLER, M., PRATCHETT, D. and CAPPER, B. S. 1977. Productivity of Africander, Tswana, Tuli and crossbred beef cattle in Botswana. *Anim. Prod.* **24:** 57–62.

(*Received* 13 *July* 1979)

34

Genetic and environmental influences on beef cattle production in Zambia. 3. Carcass characteristics of purebred and reciprocally crossbred progeny

by

W. Thorpe, D.K.R. Cruickshank and R. Thompson

Anim. Prod. 1980, 30: 245–252

GENETIC AND ENVIRONMENTAL INFLUENCES ON BEEF CATTLE PRODUCTION IN ZAMBIA

3. CARCASS CHARACTERISTICS OF PUREBRED AND RECIPROCALLY CROSSBRED CASTRATED MALES

W. THORPE AND D. K. R. CRUICKSHANK

Central Research Station, Mazabuka, Zambia

AND

R. THOMPSON

ARC Unit of Statistics, University of Edinburgh, Edinburgh EH9 3JZ

ABSTRACT

Carcass characters for 365 male castrate cattle of the Africander, Angoni, Barotse and Boran breeds, and the reciprocal crossbreds of the latter three breeds, are reported. In each of the two year-of-birth groups, different slaughter-age/management regimes were used.

For all carcass characters, except those related to size, the two sanga breeds, Africander and Barotse, were very similar, as were the two zebu breeds, Angoni and Boran. The introduced breeds, Africander and Boran, which had similar carcass weights, had heavier carcasses (+18 kg, +10%) than the indigenous Barotse and Angoni breeds. The sanga breed carcasses had less fat cover than those of the zebu breeds.

Maternal effects were not important for carcass characters and the Angoni/Barotse and Angoni/Boran crosses showed no heterosis. In the Barotse/Boran crosses, slaughter and carcass weights and eye-muscle area gave between 8% and 9·5% heterosis, and the linear carcass measurements between 2% and 3%.

It was concluded from the management comparisons that there was no economic advantage in delaying slaughter to the later of the two ages compared in the two year-of-birth groups.

INTRODUCTION

PREVIOUS papers in this series have reported the influence of breed and environmental effects on weaner production (Thorpe, Cruickshank and Thompson, 1980a) and have given live weight-for-age estimates for young stock of the indigenous breeds, Angoni and Barotse, the imported breeds Boran and Africander, and the reciprocal crossbreds of the first three breeds (Thorpe, Cruickshank and Thompson, 1980b). This paper reports the carcass characteristics of male castrates of these breed and crossbred types.

MATERIAL AND METHODS

Details of the origins of the data, routine management practices and the environment at the Central Research Station, Mazabuka, have been given by Thorpe *et al.* (1980a) and the specific management of the castrated males (steers) used in this study by Thorpe *et al.* (1980b). The data are records for purebred steers of the Africander, Angoni, Barotse and Boran breeds, and the reciprocal crossbreds of the latter three breeds, born in 1967 and 1969. All progeny were born and reared at Mazabuka apart from the purebred Africanders, which were contemporaries purchased from a commercial ranching company as calves with their dams in the 1st year, and as weaners in the 2nd year.

The two year-of-birth groups were managed differently. The 1967-born steers grazed *Hyparrhenia* veld without dry season supplementary feed and about half were chosen at random for slaughter at 3·5 years of age. The remaining steers were slaughtered at 4·5 years of age. About half the 1969-born group had supplementary feed during the latter half of their second dry season, that is after 1·5 years of age, and again for a short period prior to slaughter at 2·5 years of age. By contrast, the remaining 1969-born steers were limited to veld grazing until 90 days prior to their slaughter at 3·0 years of age, when they received the same total quantity of supplementary feed as the group slaughtered at 2·5 years of age. Of the 365 steers used in the comparison, 55 were Africanders and there were between 26 and 41 steers for each of the other genotypes.

Animals were slaughtered at a common time rather than at the same weight or condition. Seasonal rainfall distribution produces fluctuating nutritional levels from veld grazing, giving rapid weight gains during the rains and maintenance or submaintenance nutrition in the dry season. Consequently, slaughtering at constant weight or condition would confound season differences with age at slaughter. In addition, slaughtering the steers at constant ages gave estimates of comparative genotype performances which were more applicable to commercial practice.

Steers were fasted overnight, and starved live weights recorded before each group was transported 130 km to the Lusaka abattoir. Slaughtering usually took place the day after arrival, and carcasses were dressed and split in the normal manner. Wet hide and hot carcass weights, and grades were recorded. After overnight chilling, the left side was quartered and the 10th rib joint and the hump were separated. Procedures, including the specific-gravity determination of the 10th rib joint, followed those described by Ledger, Gilliver and Robb (1973). Subjective scores were given to the degree of finish and the leg shape of the chilled carcass prior to quartering, using the scale one to five, high scores being superior. The subjective scores, the linear measurements of fore-arm length, carcass width and length, 'eye muscle' A and B, and the average of the fat depths, C, D and E, were made by the same recorder according to the definitions and standards laid down by the Agricultural Research Council (1965). The characters of killing-out, forequarter, and hump percentage were based on hot carcass weights, and killing-out and hide percentages on starved live weight.

The data were analysed by least squares using a model fitting the effects block (year-of-birth/slaughter age), period of birth, genotype, the interaction of block with genotype and a regression on date of birth. The dates of birth

of the Africander steers were not known. Consequently the Africander data included variation due to date of birth and period of birth, and any variation resulting from the different pre-weaning environment. Heterosis was estimated as the difference between the reciprocal crossbred mean and the midparent breed mean.

RESULTS

Block and genotype effects were statistically significant sources of variation for most characters, and few interactions reached significance. Tables 1 and 2 give the least squares means for the genotypes and the heterosis estimates. These are expressed as percentages of the midparent mean.

Purebred comparisons. For all carcass characteristics, except those related to size, the two sanga breeds, Africander and Barotse, were very similar, as were the two zebu breeds, Angoni and Boran. Africander and Barotse steers had similar killing-out percentages, carcass shapes including hump percentages, and carcass compositions, but the imported breed, Africander, had an 18 kg (+10%) heavier carcass than the indigenous Barotse. Similarly,

TABLE 1

Means and heterosis estimates for weight and linear carcass characters

	Starved live weight at slaughter (kg)	Carcass weight (kg)	Fore-arm length (cm)	Carcass width (cm)	Carcass length (cm)
Purebred					
Africander	392·5	203·8a	39·2	47·2	119·8
Angoni	336·0	182·4b	35·3	42·8	112·3
Barotse	360·8a	185·5b	36·8a	44·4a	116·6a
Boran	370·1a	200·4a	36·8a	44·3a	116·9a
Approx. s.e. of difference between purebred means	8·04	4·70	0·37	0·38	0·91
Reciprocal crossbreds					
Angoni × Barotse	358·8a	192·2a	36·0a	43·6a	116·0a
Barotse × Angoni	352·5a	186·1a	36·3a	43·9a	114·8a
Heterosis					
units	7·2	5·2	0·1	0·1	0·9
(%)	2·1	2·8	0·3	0·3	0·8
Angoni × Boran	361·5a	198·4a	36·1a	43·4a	116·0a
Boran × Angoni	364·3a	197·9a	36·4a	43·8a	116·0a
Heterosis					
units	9·8	6·7	0·2	0·0	1·4
(%)	2·8	3·5	0·6	0·1	1·2
Barotse × Boran	390·4a	209·2a	38·2a	45·2a	119·7a
Boran × Barotse	397·6a	213·5a	37·5a	45·2a	120·1a
Heterosis					
units	28·5**	18·4***	1·0**	0·8*	3·1**
(%)	7·8	9·5	2·8	1·9	2·7
Approx. s.e. of difference between reciprocal crossbred means	9·14	5·34	0·42	0·44	1·04
Approx. s.e. of heterosis estimates	8·65	5·05	0·39	0·41	0·98

a, b. In this and following Table, means within purebred and reciprocal crossbred classes followed by the same letter do not differ significantly ($P > 0·05$).

TABLE 2

Means and heterosis estimates for carcass composition characters

	Killing-out percentage	Hide percentage	Forequarter percentage	Hump percentage	Leg shape (points)	Carcass finish (points)	Av. fat depth (mm)	'Eye muscle' A × B (mm^2)	Specific gravity, 10th rib joint	Percentage of carcass†		
										Fat	Lean	Bone
Purebred												
Africander	51·8a	9·3a	47·4a	0·41a	2·8a	1·8a	6·0a	64·9a	1·079a	14·8	64·8	20·4
Angoni	54·3b	8·8b	47·9b	0·86b	3·0a	2·1bc	8·9b	64·4a	1·063b	19·4	62·6	18·0
Barotse	51·5a	8·9b	47·3a	0·33a	2·8a	1·9ab	6·5a	62·0a	1·086a	12·7	65·7	21·6
Boran	54·2b	9·6a	47·6ab	0·77b	3·0a	2·3c	8·0b	62·4a	1·066ab	18·6	63·0	18·5
Approximate s.e. of difference between purebred means	0·39	0·18	0·25	0·046	0·13	0·12	0·60	2·32	0·0067	—	—	—
Reciprocal crossbreds												
Angoni × Barotse	53·6a	8·7a	47·1a	0·47a	3·0	2·1	7·5a	65·2a	1·067a	18·3	63·1	18·6
Barotse × Angoni	52·8a	8·9a	47·1a	0·53a	2·5	1·7	8·0a	61·6a	1·068a	18·0	63·3	18·7
Heterosis												
units	0·3	−0·1	−0·5	−0·09	−0·1	−0·1	0·0	0·2	−0·007			
(%)	0·6	−0·6	−1·5	−16·0	−5·2	−5·0	0·6	0·3	−0·6			
Angoni × Boran	54·9a	9·2a	47·6a	0·74a	3·2a	2·4a	10·3	65·3a	1·064a	19·1	62·7	18·2
Boran × Angoni	54·3a	9·2a	47·6a	0·75a	3·1a	2·4a	8·8	66·3a	1·062a	19·7	62·4	17·8
Heterosis												
units	0·3	0·0	−0·1	−0·07	0·1	0·2	1·1	2·4	0·003			
(%)	0·6	0·0	−0·3	−8·6	5·0	9·1	13·0	3·8	0·3			
Barotse × Boran	53·6a	9·4a	47·1a	0·48a	2·8	2·1a	8·5a	67·9a	1·069a	17·7	63·4	18·9
Boran × Barotse	53·6a	9·1a	47·4a	0·54a	3·1	2·3a	7·4a	67·6a	1·067a	18·3	63·1	18·6
Heterosis												
units	0·7	0·0	−0·2	−0·04	0·0	0·1	0·7	5·5*	−0·008			
(%)	1·4	0·0	−0·4	−7·3	1·7	4·8	9·7	8·9	−0·7			
Approximate s.e. of difference between reciprocal crossbred means	0·45	0·20	0·28	0·053	0·15	0·14	0·68	2·64	0·0076	—	—	—
Approximate s.e. of heterosis estimates	0·42	0·19	0·27	0·050	0·14	0·13	0·64	2·49	0·0072			

† Tissue percentages derived from the table in Ledger *et al.* (1973).

the imported zebu breed, Boran, had an 18 kg (+10%) heavier carcass than the indigenous Angoni, and similar killing-out percentage, carcass shape and composition. Compared with the Africander, the Boran had a lighter slaughter weight, a similar carcass weight and a larger hump percentage. Boran carcasses were also shorter, narrower and had more fat cover. The carcass weights of the Angoni and Barotse breeds did not differ significantly.

Tissue percentages, estimated by applying the equation derived by Ledger *et al.* (1973) to the specific gravity results, indicated that the sanga breeds had about 5 to 6% less fat tissue and about 2 to 3% more lean tissue and bone than the zebu breeds (Table 2). The lower fat percentages of the sanga breed carcasses were reflected in their lower carcass finish scores, a subjective measure used in carcass grading. As a result, some 10 to 15% of sanga carcasses did not attain top grade. This represented a substantial economic penalty. All zebu carcasses were top graded.

Reciprocal crossbred comparisons and heterosis effects. Means and heterosis estimates for the reciprocal crossbreds are presented in Tables 1 and 2. The Africander breed did not contribute to the crossbred information.

TABLE 3

Management effect means for carcass characters

Year of birth	Slaughter age (years)	Live weight (kg)	Carcass weight (kg)	Killing-out (%)	Hide (%)	Fore-quarter (%)	Hump (%)
1967	3·5	371·4	195·2	52·6	8·7	46·2	0·46
	4·5	409·5	219·7	53·6	9·0	48·3	0·69
1969	2·5	347·6a	185·1a	53·2	9·8	47·6a	0·55
	3·0	350·7a	189·4a	54·0	9·0	47·5a	0·61
s.e. of difference between means		5·16	2·99	0·26	0·12	0·16	0·031

Year of birth	Slaughter age (years)	Fore-arm length (cm)	Carcass width (cm)	Carcass length (cm)	Leg shape (points)	Carcass finish (points)	Fat depth (mm)
1967	3·5	36·8	44·4	117·2	3·1	1·9a	7·7a
	4·5	38·5	46·5	121·5	3·4	1·8a	8·4a
1969	2·5	36·3a	43·7a	114·9a	2·7	2·3	6·9
	3·0	36·4a	43·5a	114·2a	2·5	2·5	8·5
s.e. of difference between means		0·24	0·25	0·60	0·08	0·08	0·39

Year of birth	Slaughter age (years)	'Eye muscle' A × B (mm^2)	Specific gravity, 10th rib joint	Fat (%)†	Lean (%)†	Bone (%)†
1967	3·5	65·2	1·070a	17·4	63·6	19·1
	4·5	68·8	1·066a	18·6	63·0	18·5
1969	2·5	63·4a	1·082	13·9	65·2	20·9
	3·0	61·6a	1·061	20·0	62·2	17·7
s.e. of difference between means		1·53	0·0045	—	—	—

a. Slaughter age means within a year-of-birth followed by the same letter do not differ significantly.

† As in Table 2.

Tests for differences between reciprocal crossbreds were generally non-significant indicating the absence of any maternal effect on carcass characters. Heterosis estimates were also generally non-significant, with most of the Angoni/Barotse and Angoni/Boran estimates close to zero and none statistically significant. However, significant heterosis was shown by the Barotse/Boran reciprocals for characters related to size. Slaughter and carcass weight, and 'eye muscle' A × B, gave between 8% and 9·5% heterosis, and the linear carcass measurements between 2% and 3% heterosis. For these characters the performance of the reciprocals exceeded the performance of the better parent breed.

Management comparisons. Estimates of the effects of management are presented in Table 3. Within the 1967-born group, the steers slaughtered at 4·5 years of age achieved a significantly higher slaughter weight, higher killing-out percentage and heavier carcass weight (+24·4 kg, +12·6%) than the steers slaughtered at 3·5 years. Linear carcass measurements reflected this size difference and measures of fatness tended to be higher in the 4·5 year group. At both ages, about 24% of the purebred sanga carcasses failed to achieve top grade. There were no significant size differences between the 2·5 and 3·0 years slaughter age groups born in 1969. Differences in carcass composition were suggested by the thicker fat depths and lower joint specific gravity of the 3·0-year-old group. Derived tissue percentages indicated 6% more fat tissue in the 3·0-year carcasses. The greater fat thickness was reflected in a better carcass finish score, and all 3·0-year carcasses were top graded. At 2·5 years all zebu carcasses and 86% of sanga carcasses achieved top grade.

DISCUSSION

The scope for genetic improvement of beef carcass characters by inter-breed selection and crossbreeding in Zambia has been clearly demonstrated in these results. Among the purebreds, the imported zebu and sanga breeds, Boran and Africander, produced heavier carcasses than their equivalent indigenous breed types, Angoni and Barotse. Unlike the Boran, Africander carcasses lacked finish when produced under the conditions imposed in the experiment. Poorer commercial grades resulted, giving the Boran an economic advantage over all purebreds. More intensive management would seem necessary to ensure adequate fat cover for the sanga breed carcass at these slaughter ages, whereas the zebu breeds attained top commercial grades without supplementary feeding.

Preston and Willis (1974) concluded that heterosis is generally limited to carcass characters associated with size. The present results support that conclusion, as do recent American reports (Urick, Knapp, Hiner, Pahnish, Brinks and Blackwell, 1974; Long and Gregory, 1975).

The level of heterosis which may reflect the degree of breed inter-relationship is less predictable. The lack of heterosis in the Angoni/Boran crossbreds was not surprising in view of the similarity of the parent breed types (Mason and Maule, 1960). However, the generality of the relationship between level of heterosis and the degree of breed interrelationship is brought into question by the additive performance of the Angoni/Barotse crossbreds, particularly as significant heterosis was observed in the Barotse/Boran crosses. Nevertheless, the improvement shown by the crossbreds as compared to the poorer parent would give substantial increases in carcass production. This

improvement could be exploited in circumstances where the poorer breed predominates, as occurs for the Angoni in the east of Zambia and for the Barotse in the west. Basing the selection of the sire breed solely on male progeny carcass performance, the results suggest that the Boran rather than the Barotse should be recommended for crossing on the Angoni. With similar management, an improvement of some 8·5% in carcass weight could be expected. Because of the significant heterosis shown by the Boran/Barotse crosses, a 15% increase in the carcass weight would occur if Boran sires were used on the large Barotse population. These results could be expected from the live-weight performances of these genotypes (Thorpe *et al.*, 1980b), and are smaller improvements than those shown by Friesian- and Hereford-sired progeny of these breeds, which gave up to 23·5% more carcass weight than purebreds (Thorpe, Cruickshank and Thompson, 1979). Trail, Sacker and Marples (1971) found little important heterosis for carcass traits in crosses of Angus, Boran and Red Poll sire breeds, and Ankole, Boran and zebu dam breeds in Uganda, and breed of sire by breed of dam interactions were only significant for cold dressed weight and leg and chest measurements. The heterosis levels were not reported.

It is possible that level of heterosis may be positively correlated with management level (Hedrick, Krause, Lasley, Sibbit, Langford and Dyer, 1975; Long and Gregory, 1975). The low levels of heterosis in the present experiment may therefore be due to the slow live-weight gains within this environment and at these management levels (Thorpe *et al.*, 1980b).

It can be concluded from the management comparisons that there is no economic advantage in delaying slaughter to the later of the two ages compared in the two year-of-birth groups. The absence of any genotype × 'management' interaction suggests that the conclusion is applicable to all genotypes considered here. Lower slaughter ages will reduce grazing pressures and allow the expansion of either the finishing or the breeding herd with greater economic return. Exceptions may occur when carcass grade threshold values give major differences in economic return, either between breeds or between years. For example, later slaughter ages may be necessary to ensure adequate fat cover on sanga breed carcasses or to produce economic carcass weights for all genotypes following a drought year. Fat levels of the steers in the Ugandan study were also insufficient for prevailing commercial requirements and later slaughter ages were recommended (Trail *et al.*, 1971). However, as nutritional levels are dependent upon seasonal variation in grazing availability and/or feed supplementation, it can be argued that the deferment of slaughter over-emphasizes the value of carcass finish at the expense of the more economic production of carcass weight by younger stock, particularly in circumstances where meat quantity is a more important criterion than meat quality. A re-assessment of existing carcass grading systems would seem necessary to ensure that grading standards are appropriate to encourage the efficient production of lean meat, rather than the inefficient production of fat.

REFERENCES

AGRICULTURAL RESEARCH COUNCIL. 1965. *Recommended Procedures for Use in the Measurement of Beef Cattle and Carcasses.* Agricultural Research Council, London.

HEDRICK, H. B., KRAUSE, G. F., LASLEY, J. F., SIBBIT, B., LANGFORD, L. and DYER, A. J. 1975. Quantitative and qualitative carcass characteristics of straightbred and reciprocally crossed Angus, Charolais and Hereford steers. *J. Anim. Sci.* **41:** 1581–1591.

LEDGER, H. P., GILLIVER, B. and ROBB, J. M. 1973. An examination of sample joint dissection and specific gravity techniques for assessing the carcass composition of steers slaughtered in commercial abattoirs. *J. agric. Sci., Camb.* **80:** 381–392.

LONG, C. R. and GREGORY, K. E. 1975. Heterosis and management effects in carcass characters of Angus, Hereford and reciprocal cross cattle. *J. Anim. Sci.* **41:** 1572–1580.

MASON, I. L. and MAULE, J. P. 1960. *The Indigenous Livestock of Eastern and Southern Africa.* Commonwealth Agricultural Bureaux, Farnham Royal, Buckinghamshire.

PRESTON, T. R. and WILLIS, M. B. 1974. *Intensive Beef Production.* 2nd ed., pp. 157–161. Pergamon, Oxford.

THORPE, W., CRUICKSHANK, D. K. R. and THOMPSON, R. 1979. The growth and carcass characteristics of crosses of Hereford and Friesian with Angoni, Barotse and Boran cattle in Zambia. *J. agric. Sci., Camb.* **93:** 423–430.

THORPE, W., CRUICKSHANK, D. K. R. and THOMPSON, R. 1980a. Genetic and environmental influences on beef cattle production in Zambia. 1. Factors affecting weaner production from Angoni, Barotse and Boran dams. *Anim. Prod.* **30:** 217–234.

THORPE, W., CRUICKSHANK, D. K. R. and THOMPSON, R. 1980b. Genetic and environmental influences on beef cattle production in Zambia. 2. Live weights for age of purebred and reciprocally crossbred progeny. *Anim. Prod.* **30:** 235–243.

TRAIL, J. C. M., SACKER, G. D. and MARPLES, H. J. S. 1971. Crossbreeding beef cattle in western Uganda. 5. Growth and carcass evaluation of castrated males. *Anim. Prod.* **13:** 171–180.

URICK, J. J., KNAPP, B. W., HINER, R. L., PAHNISH, O. F., BRINKS, J. S. and BLACKWELL, R. L. 1974. Results from crossing beef × beef and beef × Brown Swiss: carcass quantity and quality traits. *J. Anim. Sci.* **39:** 292–302.

(*Received* 13 *July* 1979)

35

Genetic and environmental influences on beef cattle in Zambia. 4. Weaner production from purebred and reciprocally crossbred progeny

by

W. Thorpe, D.K.R. Cruickshank and R. Thompson

Anim. Prod. 1981, **33**: 165–177
© 1981 British Society of Animal Production
0003-3561/81/00110165$02·00

GENETIC AND ENVIRONMENTAL INFLUENCES ON BEEF CATTLE PRODUCTION IN ZAMBIA

4. WEANER PRODUCTION FROM PUREBRED AND RECIPROCALLY CROSSBRED DAMS

W. THORPE† AND D. K. R. CRUICKSHANK‡
Central Research Station, Mazabuka, Zambia

AND

R. THOMPSON
ARC Unit of Statistics, University of Edinburgh, Edinburgh EH9 3JZ§

ABSTRACT

Estimates of dam reproductive performance are reported for purebred Africander, Angoni, Barotse and Boran dams, and the reciprocal crossbreds of the latter three breeds, under ranching conditions in Zambia.

Mean performances based on a maximum of 1996 records collected over six parities were: calving rate, 65·6%; weaning rate, 62·8%; calf birth weight, 26·6 kg; calf weaning weight, 163·6 kg; and dam live weight at weaning, 363·0 kg.

Purebred Africander and Barotse dams had significantly poorer calving rates, 54·2 and 57·6% respectively, than the purebred dams of the Angoni and Boran breeds, 69·1 and 66·0% respectively. Reciprocal crossbred dams of the latter two breeds had the highest calving rates, 71·2 and 74·4%. Heterosis estimates for calving rate were 0·8% for Angoni/Barotse crosses, 7·8% for Angoni/Boran crosses and 10·3% for Barotse/Boran crosses and, for weaning rate, 3·7, 6·7 and 11·5% respectively. There were no differences between reciprocal crossbred dams for calving and weaning percentages. The mortality rate of calves born to crossbred dams was 4·0% and that of the calves born to purebred dams of the same breeds 4·7%. Dam status at mating had an important effect on fertility with the calving rate of dams dry at mating, 89·1%, contrasting with a rate of 39·8% for those dams that were lactating at mating and that had calved late in the season.

There were large differences between breeds for calf weights but heterosis estimates were small and non-significant. There were no differences between the progeny weights of reciprocal crossbred dams. Sex, date and period of birth, and dam status had significant effects on calf weights.

Africander dams had the heaviest, and Angoni dams the lightest, live weights at weaning, 421·1 and 320·7 kg respectively. Heterosis estimates for dam live weights were small and non-significant, and reciprocal crossbred differences were unimportant. Dams which lost most live weight during the dry season, that had extremely light live weights at the beginning of the breeding season or that gained little weight during the breeding season had significantly poorer fertility than other dams.

Overall estimates of heritability derived by dam/daughter regression were 0·20 for calving percentage, 0·76 for calf birth weight and 0·72 for calf weaning weight.

INTRODUCTION

IMPROVEMENTS in beef production through crossbreeding have generally been greater in reproductive than in growth and efficiency traits. Particular advantages have been found in utilizing first-cross dams (Koger, 1973). Consequently, an important part of the breed evaluation programme reported in this series of papers was the estimation of the comparative reproductive performances of purebred Africander, Angoni, Barotse and Boran dams, and the reciprocal crossbred dams of the latter

†Present address: Facultad de Medicina Veterinaria y Zootecnia, Universidad de Yucatan, Apartado 116D, Merida, Mexico.

‡Present address: Mukulaikwa Agricultural Station, PO Box 30613, Lusaka, Zambia.

§Address for offprints.

three breeds. The reproductive performance of these 10 genotypes over six parities is reported in this paper.

The design of the experiment allowed additive gene effects, maternal effects and heterosis to be estimated for calving and weaning rates, calf birth and weaning weights, and dam live weights. The analyses provided estimates of environmental effects for these traits and relationships were derived between subsequent calving rate, and dam live weights and live-weight changes. As reproductive and calf live-weight records were available for the dams of these breeding females (Thorpe, Cruickshank and Thompson, 1980a), it was possible to estimate dam/daughter regressions for calving rate and calf weights. Live-weight estimates from birth to 3·5 years of age for the breeding females used in the experiment have been given by Thorpe, Cruickshank and Thompson (1980b).

MATERIAL AND METHODS

Management

Females of the nine possible genotype combinations of the Angoni, Barotse and Boran breeds born at the Central Research Station, Mazabuka, in 1967 and 1969, and purchased contemporary samples of purebred Africander, provided the reproductive records for which analyses are reported. Management prior to mating has been described in detail by Thorpe *et al.* (1980b). The young females (heifers) born in 1967 were reared solely on veld grazing and were mated at 3 years of age. The heifers born in 1969 received supplementary feed during their second dry season and were mated at 2 years of age.

The group born in 1967 had their first and second calf crops at Mazabuka. Their subsequent calf crops and all calf crops for the herd born in 1969 were produced at Mochipapa Research Station, Choma. At Mazabuka the herd born in 1967 grazed upland veld, and during the dry seasons of their first and second calf crops received some supplementary feed on a herd basis. Other management practices followed the routine described for the dams of these females by Thorpe *et al.* (1980a).

Management at Mochipapa was very similar to that at Mazabuka, with a calving season from late September to December and a breeding season from late December to late March. Calves were weaned on the same day, generally in late June. Throughout the experiment, dams barren for 2 consecutive years were culled. An *ad hoc* rotational grazing system was practised and standing hay was retained for dry-season grazing. In addition dams were given supplementary feed from August to December on a herd basis. The feed generally consisted of a mixture of 800 g ground snapcorn and 200 g of a protein concentration with a mineral additive per kg, and contained 100 to 120 g crude protein per kg. It was given on a scale rising from approximately 1 kg per head per day during August to 2 kg per head per day during October and November. The endpoint of the supplementation varied with the onset of the rains, which determined the availability of the new season's herbage.

Environment

Details relating to the Central Research Station, Mazabuka have been given by Thorpe *et al.* (1980a). Mochipapa Research Station is at an elevation of approximately 1370 m on the Southern Plateau of Zambia at latitude 16°52′ S and longitude 27°06′ E. The station has unimproved acid sandveld grazing with large areas of seasonally waterlogged depressions called dambos (Astle, 1966). The grazing is typical of that supporting a large part of Zambia's cattle population and was expected to be less productive than that at Mazabuka. Between 1972 and 1977, the duration of the experiment at Mochipapa, mean annual rainfall was 738 mm, with a range from 337 to 1015 mm.

Sire breeds

Progeny were sired by different breeds in each parity (Table 1). The last records available were the sixth and fifth parities for the 1967- and 1969-born groups respectively. These were the matings with Barotse sires in 1976. This mating will have introduced some small bias to the estimates of maternal effects for the Barotse crossbred dams. The design of the experiment meant that it was not possible to separate the effects of dam year of birth, progeny sire breed and breeding season; these effects were combined in the analyses and reported as a dam parity code.

Description of data and statistical analyses

The traits analysed were calving and weaning

TABLE 1

Parity code related to dam year of birth, breeding season and sire breed

	Dams born in 1967		Dams born in 1969	
Parity code	Breeding season	Sire breed	Breeding season	Sire breed
1	1971	Hereford	1972	Hereford
2	1972	South Devon	1973	South Devon
3	1973	Friesian	1974	Friesian
4	1974	Bhagnari	1975	Bhagnari
5	1975	Sindhi	1976	Barotse
6	1976	Barotse		

percentages, calculated as the proportion of live calves born and weaned to dams mated each season, calf birth and weaning weights; and dam live weights at the beginning and end of the breeding season, at the autumn peak and at weaning. The dam live weights have been described by Thorpe *et al.* (1980a). All dams except the 1969-born herd in their first parity were weighed at weaning and, from October 1975, dams were weighed at 4-week intervals. Both herds were weighed monthly while at Mazabuka.

In the hierarchical analyses of variance for calving and weaning percentages, effects were fitted for dam status at mating (dry; lactating <294 days, 294 to 314 days, >314 days), dam previous parous state (dry, lactating), dam parity code (1, 2, ... 6), dam year of birth (1967, 1969) $\times$ parity, Africander, sire breed (Angoni, Barotse, Boran), dam breed (Angoni, Barotse, Boran) and sire breed $\times$ dam breed interaction. The division of lactating status at mating grouped dams which had calved early (i.e. before 21 October (<294 days counting from 1 January)), in mid season (i.e. between 12 October and 10 November (294–314 days)) and which had calved late (i.e. after 10 November (>314) days)). The number of reproductive records for each genotype and environmental class are given in Tables 2 and 3 respectively.

The model for the dam live weight analyses included, in addition to the above effects, dam status at next calving (calved, not calved). Tables 5 and 6 give the number of records used in the dam live-weight analyses.

Hierarchical analyses of variance were used to estimate relationships between (a) subsequent calving rate and live weight at the beginning of the breeding season, and (b) subsequent calving rate and breeding season live-weight change. To derive these relationships, calving percentage was analysed using the model given above with, in addition, the effects of live weight (grouped in 11 classes with 12-kg intervals) and live-weight change (grouped in 8 classes with 10 kg intervals). The live weight and live-weight change classes had a minimum of 55 and 75 records respectively. The relationship between the subsequent calving percentages of lactating dams and the live-weight change of the dams from the autumn peak to the mean of the breeding season weights was also derived. Records were available for the 1967-born herd in their second, third and sixth parities and for the fifth parity of the 1969-born herd. There were 499 live-weight change records with a minimum of 48 records per class live-weight change.

In the analyses of the calf weights, effects were fitted for sex (male, female), period of birth (early calved: <294 days; mid calved: 294 to 314 days; late calved: >314 days), dam previous parous state (dry, lactating), dam parity, dam year of birth $\times$ parity interaction, Africander sire breed of dam, dam breed of dam, sire breed of dam $\times$ dam breed of dam interaction and a covariate for date of birth. The period of birth effect estimated the influence of season. Calving began in late September during the dry season and ended in late December when the rains were usually established. The onset of effective rain generally occurred in mid November, i.e. after day 314.

Dam/daughter regressions were estimated for the dam traits calving percentage, and calf birth and weaning weights, from records of purebred Angoni, Barotse and Boran dams (Thorpe *et al.*, 1980a) and from the records of the present cows—which are their purebred and crossbred daughters. In the estimation of the regressions for

calf birth and weaning weights the model included the effects given for the analysis of these calf weights. In addition, regressions for date of birth and dam birth or weaning weights were included as appropriate. In the Angoni, Barotse, Boran and pooled dam breed analyses there were 90, 92, 90 and 272 females respectively. Table 7 gives the total number of records used in the calf weight regression estimates.

In the estimation of calving percentage regressions, only records of dams with at least three exposures to the bull were used. Table 7 gives the total number of records in each analysis. There were 89, 93, 88 and 270 females respectively in the Angoni, Barotse, Boran and pooled breed analyses.

RESULTS

Calving and weaning percentages

Genotype means and heterosis estimates are presented in Table 2. Mean calving rate was 65·6%. The zebu breeds, Angoni and Boran, had the highest purebred calving rates, 69·1 and 66·0% respectively, and their reciprocal crosses had the highest calving rates of all genotypes, 71·2 and 74·4%. Purebred dams of the sanga breeds, Africander and Barotse, had the lowest calving rates, 54·2 and 57·6% respectively. Heterosis in the Angoni, Barotse and Boran breeds was lowest in the Angoni/Barotse crosses (0·8%), intermediate in the Angoni/Boran crosses (7·8%), and highest in the Barotse/Boran crosses

TABLE 2

Genotype means and heterosis estimates for calving and weaning percentages and calf birth and weaning weights

Dam genotype	No. of dam records	Calving percentage	Weaning percentage	Calf birth weight (kg)	Calf weaning weight (kg)
Overall mean	1996	65·6	62·8	26·6	163·6
Africander	412	54·2[a]†	51·4[a]	31·1	174·0[a]
Angoni	200	69·1[c]	65·1[c]	23·9	149.1
Barotse	136	57·6[ab]	53·8[ab]	27·1[a]	162·6[b]
Boran	153	66·0[bc]	64·5[bc]	26·7[a]	168·9[ab]
Av. s.e. of differences between purebred means		6·32	6·00	0·64	4·22
Angoni × Barotse	190	67·3[a]	64·4[a]	25·4[a]	156·7[a]
Barotse × Angoni	174	60·4[a]	59·1[a]	25·7[a]	159·3[a]
Heterosis					
Units		0·5	2·3	0·1	2·1
%		0·8	3·9	0·3	1·4
Angoni × Boran	190	74·4[a]	70·8[a]	25·3[a]	161·9[a]
Boran × Angoni	203	71·2[a]	67·4[a]	26·0[a]	158·7[a]
Heterosis					
Units		5·2	4·3	0·9	1·3
%		7·8	6·7	3·5	0·8
Barotse × Boran	163	68·4[a]	67·1[a]	27·5[a]	175·9[a]
Boran × Barotse	175	67·8[a]	64·7[a]	27·3[a]	169·2[a]
Heterosis					
Units		6·3	6·8	0·5	6·8
%		10·3	11·5	1·8	4·1
Av. s.e. of difference between reciprocal crossbred means		6·40	6·07	0·62	4·09
Av. s.e. of heterosis estimates		6·60	6·27	0·66	4·31

†a, b, c: Those means within purebred and within reciprocal crossbred classes followed by the same letter do not differ significantly ($P > 0·05$).

(10·3%). None of the heterosis estimates were significantly different from zero ($P>0·05$) and there were no significant differences between the reciprocal crossbreds.

Results for weaning percentage were very similar to those for calving percentage and there were no important rank changes between genotypes. The heterosis estimates, 3·9, 6·7 and 11·5% for the Angoni/Barotse, Angoni/Boran and Barotse/Boran crosses respectively, were not significantly different from zero and there were no significant differences between the reciprocal crossbreds.

The mean weaning rate, 62·8%, gave an overall estimate of calf pre-weaning mortality of 4·3%. The mortality rate of calves born to crossbred dams (4·0%) was marginally lower than the rate (4·7%) for the calves born to the purebred dams of the Angoni, Barotse and Boran breeds. The mortality rate for progeny of Africander dams was similar to that for progeny born to other purebred dams.

TABLE 3

Effect of dam parous state on calving percentage

	No. of dam records	Calving percentage
Overall mean	1996	65·6
Dam status at mating		
Dry	860	89·1
Lactating		
Early calved	409	68·9[a]†
Mid calved	377	64·7[a]
Late calved	350	39·8
Dam previous parous state		
Dry	1138	65·4[a]
Lactating	858	65·9[a]

†In this and following Tables, means within a class followed by the same letter do not differ significantly ($P>0·05$).

Dam status at mating had an important effect on fertility (Table 3). Dams dry at mating had the highest fertility (89·1%), which was some 20 percentage units better than the fertility of the lactating dams which had calved during the early and middle periods of the calving season. The fertility of the dams which had calved after 10 November, and which therefore had the shortest post-parturition period before the start of the breeding season, was depressed by a further 25 percentage units. Reproductive performance in the previous parity did not affect fertility.

Calf birth and weaning weights

Genotype means and heterosis estimates are presented in Table 2. At birth and weaning the progeny of purebred Angoni dams were significantly smaller than the progeny of purebred Barotse and Boran dams, the weights of which did not differ. At both weights progeny of the purebred Africander dams were significantly heavier than the progeny of the other purebred dams with the exception of the progeny of Boran dams at weaning.

Crossbred dam performances for calf birth and weaning weights were close to the mean of the performances of the parental purebreds and the heterosis estimates were not significantly different from zero. Calf weaning weights of the Angoni/Barotse and Angoni/Boran reciprocal crossbred dams were less than 2% above the mean of the parental purebreds, and in neither case were the progeny of the crossbred dams superior to those of the better parent breed. Weaning weights of the progeny of the Barotse/Boran crossbred dams were equal to or better than the weaning weights of the better parental breed, the Boran, but the differences were not significant. There were no reciprocal crossbred differences for birth or weaning weights.

Estimates of environmental effects on calf weights are presented in Table 4. At birth and weaning, male calves were approximately 5% heavier than female calves. The seasonal effect, period of birth, indicated that the calves born in the latter part of the season had lighter birth weights and the regression of calf weight on date of birth indicated that the earlier in the season a calf was born the heavier was its birth weight. The large regression of calf weaning weight on date of birth was not significant but suggested that the later-born calves were lighter at weaning. Dams which had not calved the previous season gave birth to and weaned significantly heavier calves than dams which calved in two successive seasons.

Dam live weights

Estimates of means and heterosis for the live weights of the purebred and reciprocally crossbred dams are presented in Table 5. The

TABLE 4

Estimates of environmental effects on calf birth and weaning weights

	No. of records	Calf birth weight (kg)	No. of records	Calf weaning weight (kg)
Overall mean	1297	26·6	1263	163·6
Sex				
Male	657	27·2	644	168·1
Female	640	26·0	619	159·1
s.e. of difference		0·21		1·18
Date of birth regression ± s.e.	1297	0·057 0·009	1263	−0·502 0·435
Period of birth				
Early calved	459	27·1[a]	453	163·0[a]
Mid calved	414	26·6[a]	404	163·7[a]
Late calved	424	26·0[b]	406	164·2[a]
Av. s.e. of difference		0·42		2·40
Dam previous parous state				
Dry	680	27·2	663	168·2
Lactating	617	26·0	600	159·1
s.e. of difference		0·28		1·63

TABLE 5

Genotype means and heterosis estimates for dam live weights

		Dam live weight (kg)				
		Breeding season				
Dam genotype	No. of dam records	Beginning	End	Autumn peak	No. of dam records	Dam live weight at weaning (kg)
Overall mean	1092	293·5	335·2	349·5	1388	363·0
Africander	196	323·5	375·0	393·2	260	421·1
Angoni	119	264·2	299·0	311 5	149	320·7
Barotse	78	296·3[a]	339·5[a]	356·4[a]	95	374·5[a]
Boran	86	302·8[a]	345·0[a]	358·7[a]	108	371·1[a]
Av. s.e. of differences between purebred means		6·60	7·17	7·62		6·23
Angoni × Barotse	110	281·8[a]	321·8[a]	336·1[a]	137	346·8[a]
Barotse × Angoni	99	287·4[a]	328·7[a]	342·5[a]	120	353·6[a]
Heterosis						
Units		4·4	6·0	5·4		2·6
%		1·6	1·9	1·6		0·8
Angoni × Barotse	105	286·4[a]	322·9[a]	334·1[a]	135	340·3[a]
Boran × Angoni	114	278·1[a]	314·0[a]	327·1[a]	148	339·0[a]
Heterosis						
Units		−1·3	−3·6	−4·5		−6·3
%		−0·4	−1·1	−1·3		−1·8
Barotse × Boran	86	305·3[a]	349·6[a]	364·3[a]	115	376·8[a]
Boran × Barotse	99	309·6[a]	356·9[a]	370·8[a]	121	386·2[a]
Heterosis						
Units		7·9	11·0	10·0		8·7
%		2·6	3·2	2·8		2·3
Av. s.e. of difference between reciprocal crossbred means		6·71	7·29	7·75		6·39
Av. s.e. of heterosis estimates		6·89	7·49	7·96		6·53

least fertile of the purebred dams, the Africander, had the heaviest live weights and the most fertile, the Angoni, the lightest live weights. The Africander were some 100 g/kg heavier than the Barotse and Boran dams which, in turn, were some 150 g/kg heavier than the Angoni dams.

None of the heterosis estimates for the crosses between the Angoni, Barotse and Boran breeds reached significance ($P>0{\cdot}05$) and the live weights of the reciprocal crossbreds did not differ. The live weights of the crosses of the Barotse and Boran with the Angoni were generally significantly lighter than the larger parental breed and significantly heavier than the smaller parental breed, the Angoni. The Barotse/Boran crossbred dams were not significantly heavier than either parental breed.

Table 6 gives estimates for the environmental effects on dam live weights. On average, dams gained 42 kg live weight (140 g/kg) during the breeding season and a further 14 kg before reaching their peak live weight in autumn, a total live-weight gain of some 190 g/kg. The peak live weight was reached in May, some 2 months after the end of the breeding season.

Dams which were lactating when weighed were significantly lighter at all stages than dry dams. At the beginning of the breeding season the difference was less than 20 kg and increased to over 40 kg by the autumn and at weaning. Dams which had been dry the previous season began the breeding season 19 kg (70 g/kg) heavier than the previously lactating dams, a difference reduced to 4 kg at weaning 6 months later. Live weights at the different parities showed the expected increase in dam live weight with age.

Relationships between calving percentage, and dam live weights and live-weight changes

The pregnancy status results presented in Table 6 are estimates of the influence of dam live weight on conception rate as measured by success or failure to calve in the following season. Dams which subsequently calved were 5·3 kg heavier at the beginning of the breeding season than dams which did not calve. The difference increased to 16·4 kg at the autumn peak and the dams calving subsequently were 17·8 kg heavier at weaning. Each of these differences was significant.

Figure 1 shows the relationship between the range of dam live weights at the beginning of the breeding season and subsequent calving rate. Dams in the lightest live-weight class, ≤ 232 kg, had a significantly lower calving rate than dams

TABLE 6

Estimates of environmental effects on dam live weights

	Dam live weight (kg)					
	Breeding season					
	No. of records	Beginning	End	Autumn peak	No. of records	Dam live weight at weaning (kg)
Overall mean	1092	293·5	335·2	349·5	1388	363·0
Pregnancy status						
Not pregnant	326	290·9	331·0	341·3	452	354·1
Pregnant	766	296·2	339·5	357·7	936	371·9
s.e. of difference		2·22	2·37	2·50		2·35
Dam status at mating						
Dry	594	307·3	355·9	381·3	406	399·5
Lactating						
Early calved	175	281·6	323·9[a]	335·2[a]	312	351·8[a]
Mid calved	178	290·0[a]	330·6[a]	343·6[b]	334	351·3[a]
Late calved	145	295·2[a]	330·6[a]	337·7[ab]	326	349·4[a]
Av. s.e. of difference		3·74	4·10	4·32		3·28
Dam previous parous state						
Dry	759	302·8	341·6	354·2	530	364·9[a]
Lactating	333	284·2	328·9	344·8	858	361·1[a]
s.e. of difference		3·57	3·83	4·03		3·08

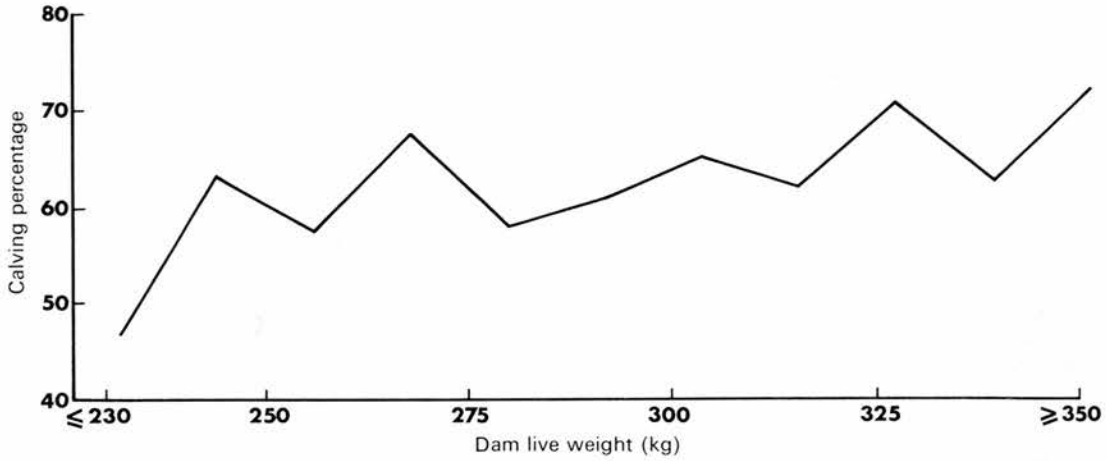


FIG. 1. Relationship between subsequent calving rate and dam live weight at the beginning of the breeding season.

in most heavier live-weight classes. The live weights of dams in these heavier classes ranged from 244 kg to 312 kg and their calving rates varied by less than 15 percentage units, differences which were generally non-significant ($P > 0{\cdot}05$).

There was a positive relationship between live-weight change during the breeding season and subsequent calving rate (Figure 2). Dams which gained 10 kg or less during the breeding season calved at 41%, whereas dams which gained 40 kg calved at 66%. The highest calving rate, 80%, was achieved by those dams which gained the most live weight, $\geqq 80$ kg.

In this population, dams which were lactating at mating had a subsequent calving rate which was, on average, 31 percentage units lower than that of dams which were dry at mating (Table 3). Figure 3 shows the relationship between the subsequent calving rate of these lactating dams and their live-weight change from the autumn peak to the mean of the breeding season weights. The mean live-weight change for the 499 lactating dams was $-47{\cdot}8$ kg and the mean

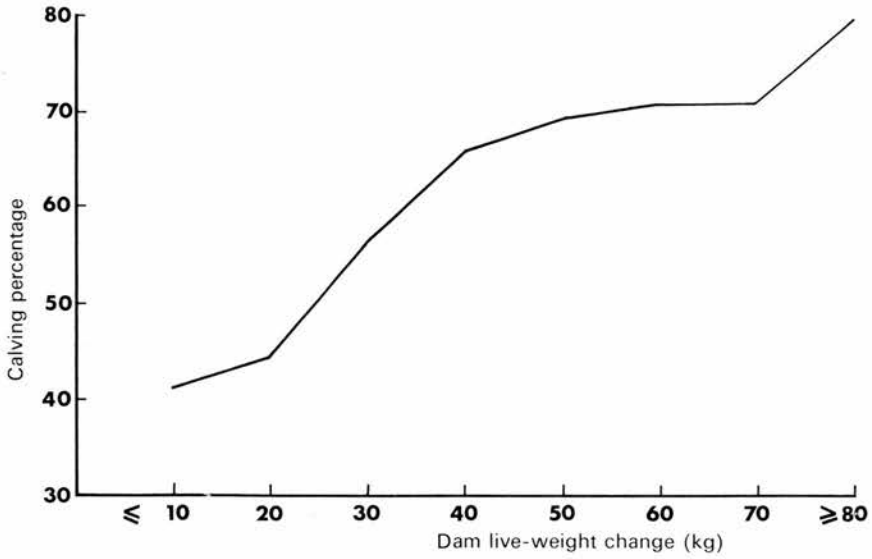


FIG. 2. Relationship between subsequent calving rate and dam live-weight change during the breeding season.

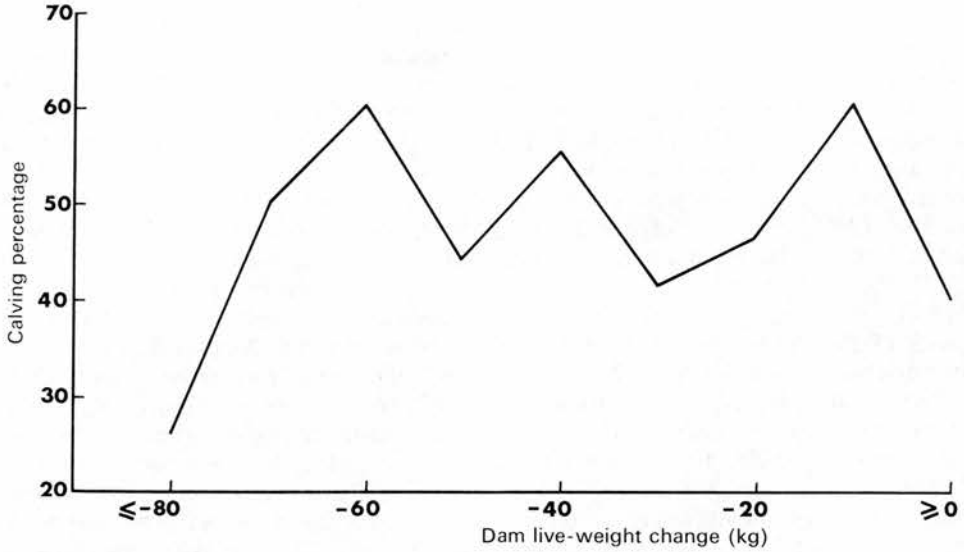


FIG. 3. Relationship between subsequent calving rate and live-weight change from autumn peak to mid-mating of lactating dams.

calving rate 47·3%. Dams which lost the most live weight, ≧80 kg, calved at a significantly lower rate, 26%, than all other dams. Calving rates of the remaining live-weight change classes ranged from 42 to 60%, differences which were not significant.

Estimates of dam/daughter regressions

Table 7 presents heritabilities for calving percentage and calf birth and weaning weights estimated as twice the dam/daughter regression. The estimates were low to moderate for calving percentage, ranging from 0·08 for Boran dams to 0·38 for Angoni dams, with an overall estimate of 0·20. Estimates for both calf weights were high, exceeding 0·70 for the pooled population and 0·50 for each of the breed populations.

DISCUSSION

Maternal heterosis estimates for zebu × zebu and zebu × sanga crosses do not appear to have been reported previously. Although the estimates derived for calving and weaning percentages, and calf birth and weaning weights, of the Angoni, Barotse and Boran breeds were not significant, they were consistently positive. If these estimates are combined in indices of dam productivity, they indicate the possibility of important improvements in weaner production. The product of weaning rate and weaning weight, the weight of weaner produced per dam exposed per year, shows 5·6, 7·6 and 15·8% heterosis for the Angoni/Barotse, Angoni/Boran and Boran/Barotse crosses respectively. These estimates can be compared with the average

TABLE 7

Estimates of heritabilities (h^2) *for calving percentage and calf birth and weaning weights by dam/daughter regression*

	Calving percentage			Calf birth weight			Calf weaning weight	
Dam breed	No.	h^2	s.e.	No.	h^2	s.e.	h^2	s.e.
Angoni	504	0·38	0·36	327	0·86	0·17	0·80	0·22
Barotse	526	0·28	0·36	316	0·58	0·14	0·91	0·19
Boran	494	0·08	0·30	337	0·79	0·12	0·50	0·21
Pooled	1524	0·20	0·18	980	0·76	0·09	0·72	0·12

values of 12·5% for *Bos taurus* breed crosses and 46·3% for *Bos indicus* × *Bos taurus* crosses derived by Koger (1973) from reports in the literature. When dam live weight at weaning is taken into account, the estimate of productivity, weight of weaner produced per 100 kg dam live weight exposed per year, gives heterosis estimates of 3·7, 9·4 and 12·9% for the Angoni/Barotse, Angoni/Boran and Boran/Barotse crosses respectively.

The heterosis estimates are closer to those among crosses of *Bos taurus* breeds than to the larger estimates shown by crosses of *Bos indicus* with *Bos taurus*, and it seems reasonable to expect this. Hence, to utilize heterosis fully, *Bos taurus* crossbreeding should be considered in Zambia. Moreover, a crossbreeding system including *Bos taurus* sires should make better use of complementary traits. Thorpe, Cruickshank and Thompson (1979) have shown that *Bos taurus*-sired F_1 progeny had relatively greater live and carcass weight advantages over purebred Angoni, Barotse and Boran progeny than did the crossbred progeny of these breeds (Thorpe *et al.*, 1980b; Thorpe, Cruickshank and Thompson, 1980c). Preliminary reproductive results suggest that Hereford-sired F_1 dams are competitive with the best of the F_1 crosses reported here (Thorpe *et al.*, 1979). Theoretical and practical considerations suggest that rotational crossbreeding systems make best use of breed resources and the results of Crockett, Koger and Franke (1978a and b) support the expectation that rotations with two breeds of sire can be effective in maintaining significant levels of heterosis in most productive traits. It therefore follows that, if optimum crossbreeding systems are to be developed to utilize Zambia's breed resources, comparisons are required between rotational crosses of *Bos taurus* breeds, and the local *Bos indicus* and sanga breeds.

Heterosis estimates for the individual traits calving percentage and calf pre-weaning survival were similar to the values for crosses among *Bos taurus* breeds reported in several American experiments, for example the study of crosses of Hereford, Angus and Shorthorn reported by Cundiff, Gregory and Koch (1974a). However, in contrast to the results in the American reports, maternal performance in the present study, as expressed by calf weaning weight, showed little heterosis except in the Barotse/Boran cows.

The contribution of heterosis to productive efficiency, when efficiency is estimated by the annual production per 100 kg dam live weight, will be influenced by the level of heterosis shown by dam live weight throughout the dam's productive life. Smith, Fitzhugh, Cundiff, Cartwright and Gregory (1976) have shown that heterosis in Hereford, Angus and Shorthorn dam live weight declined from 7·3% at 396 days to 2·5% at maturity. In the present experiment, heterosis estimates (at 3·5 years of age) for the live weights of Angoni/Barotse, Angoni/Boran and Barotse/Boran crosses were 3·8, 1·4 and 4·7% respectively (Thorpe *et al.*, 1980b). On the other hand, heterosis estimates for the dam live weights of these crosses (Table 5) were consistently lower. For example, at weaning heterosis values were 0·8, −1·8 and 2·3% respectively. There is therefore some support for the suggestion made by Smith *et al.* (1976) that the effect of heterosis on early growth rate is greater than the accompanying increase in mature weight.

Koch (1972) has reviewed several studies illustrating reciprocal crossbred differences for dam traits. More recently, two papers (Cundiff *et al.*, 1974a; Cundiff, Gregory, Schwulst and Koch, 1974b) have reported significant differences between Hereford/Angus reciprocals for fertility, milk production and pre-weaning traits. However, McDonald and Turner (1972) did not find significant differences in maternal performance between Hereford/Angus reciprocals nor in pre-weaning growth traits for the reciprocal crosses of Hereford, Angus, Brahman and Brangus. In the present study there was no evidence of reciprocal crossbred differences for maternal performance.

The poor fertility of the Africander dams resulted in their having the lowest productivity of all dam types in the comparison. In Botswana the Africander had a poorer fertility than other local breeds (Trail, Buck, Light, Rennie, Rutherford, Miller, Pratchett and Capper, 1977). Factors contributing to the low calving rate are likely to have included a protracted *post partum* anoestrus, the incidence of 'silent' oestrus and a longer gestation interval (Harwin, Lamb and Bisschop, 1967; Rakha, Igboeli and King, 1971; Holness, Hopley and Hale, 1978). In a previous Zambian study, Rakha *et al.* (1971) found that Africander dams had a mean calving interval

which was 46 days longer than that of Angoni dams.

In Botswana, purebred Africander progeny had significantly higher mortality rates to 1 and 2 years of age than the purebred progeny of other local breeds (Trail *et al.*, 1977). Similarly, Rakha *et al.* (1971) found pre-weaning mortality rates of 3·8% and 8·9% for purebred Angoni and Africander progeny respectively. However, in the present study, crossbred progeny of Africander dams did not have higher mortality rates than the progeny of other dams. The different Angoni, Barotse and Boran breed effects on pre-weaning mortality reported by Thorpe *et al.* (1980a) were not repeated in this experiment, in which the majority of the progeny were sired by an unrelated breed.

Not only the progeny mortality rates but also the relative fertilities of these latter three breeds differed in this study from those reported at Mazabuka Research Station by Thorpe *et al.* (1980a). These differences suggested that genotype × environment interactions may be important for the dam reproductive performance of these breeds and that breed comparisons should be conducted in the ecological region for which breed recommendations are required. Dams of the Barotse breed, which originates from an area which includes the Zambezi River floodplain, seem particularly disadvantaged by an acid sand veld environment, whereas the Boran breed, which originates from the arid Kenya-Ethiopia border region, appears relatively well adapted to the harsher Mochipapa environment.

The better fertility of dams which were dry at mating compared with those lactating (Table 3) is consistent with most reports in the literature. Comparable results, including the effects of stage of lactation at mating, have been reported from Botswana by Buck, Light, Rutherford, Miller, Rennie, Pratchett, Capper and Trail (1976). When breeding is restricted to the period from December to March, dams calving during the latter part of the calving season (after 10 November) have less time available to resume normal reproductive function than dams calving earlier. The period between parturition and first service for Africander and Angoni dams has been estimated as 66·9 (s.e. 20·9) and 79·9 (s.e. 19·2) days respectively (Rakha *et al.*, 1971). As the live weights of the dams in the stage of lactation at mating classes did not differ by the end of the breeding season (Table 6), it seems that the *post partum* period and its associated lactation stress exerted greater influence on fertility than any variation due to live weight. On the other hand, the fertility of purebred Barotse and Boran dams was not significantly affected by status at mating in the more favourable Mazabuka environment (Thorpe *et al.*, 1980a). In that study the significant effect shown by Angoni dams was small compared to the effect presented in Table 3. From these results it is apparent that live weight is only an approximate indicator of nutritional status in the complex interaction of nutrition, lactation and reproductive performance, and detailed studies are required of nutritional-hormonal relationships. Some indication that these relationships may differ between breeds has been provided by the work of Holness *et al.* (1978). Estimates of the relationship between condition score and reproductive performance may also prove useful.

At extremely low live weights, or when large live-weight losses occurred, fertility was depressed. Those lactating dams losing the most live weight between the autumn peak and their mid-mating weight had the poorest calving rates in the following season. A similar relationship has been reported by Richardson, Oliver and Clarke (1975) and, in common with the finding of Buck *et al.* (1976), those dams starting the breeding season with the lightest live weights had the poorest calving rates. In the Botswana study (Buck *et al.*, 1976), and in these results, dam live-weight gain during the breeding season had a positive relationship with fertility. Separation of vulnerable groups and provision of supplementary feed to increase live weight or alleviate live-weight loss is likely to give benefits in fertility. Similarly, the feeding of supplementary protein in the dry season has been shown to improve fertility (Ward, 1968; Richardson *et al.*, 1975; Capper, Pratchett, Rennie, Light, Rutherford, Miller, Buck and Trail, 1977), although Capper *et al.* (1977) found no significant weight change associated with the improvement in fertility shown by lactating dams. Reliance on live-weight records for the determination of supplementary feed levels may therefore result in uneconomic levels of feeding, particularly for dams with average and above average live weights (Figure 1) and live-weight changes (Figure 3).

Environmental effects on calf weights were relatively small compared with their effects on fertility and were generally consistent with the majority of other reports (Preston and Willis, 1974). On the other hand, the heritabilities of the calf weights, estimated as twice the dam/daughter regression, were higher than the average values given by Preston and Willis (1974), and the estimates reported by Tonn (1975) for Boran progeny in Kenya and by Trail, Sacker and Fisher (1971) for progeny of breeds including the Boran. Similarly, the indications are that fertility in the Angoni and Barotse breeds was moderately heritable, although it is noted that the estimates were subject to considerable sampling variation (Table 6). Some other moderate estimates of calving rate heritability have been reported (Brinks, Clark, Kieffer and Urick, 1964; Deese and Koger, 1967) but fertility is generally considered to have low heritability ($<0{\cdot}10$). Further data are required to substantiate the calving rate heritability estimate. If confirmed, dam selection may prove a feasible method of improving production.

ACKNOWLEDGEMENTS

The authors thank the Director of Agriculture for the provision of experimental facilities and acknowledge the technical assistance of the Animal Husbandry Research staff at Mazabuka and at Mochipapa. Particular thanks are due to the Officer in Charge, Mochipapa, and to G. Simakando, A. Chisani, M. Mooya and J. Phiri. Thanks are also due to Professor J. H. D. Prescott, Edinburgh School of Agriculture, Professor D. J. Finney, ARC Unit of Statistics, Dr G. Wiener and Mr I. L. Mason. The UK Overseas Development Administration financed the reporting of the results through research scheme R3390.

REFERENCES

ASTLE, W. L. 1966. Edaphic grasslands in Zambia. *Proc. int. Grassld Congr., Brazil*, Vol. 1, pp. 363–373.

BRINKS, J. S., CLARK, R. T., KIEFFER, N. M. and URICK, J. J. 1964. Predicting producing ability in range Hereford cows. *J. Anim. Sci.* **23:** 593 (Abstr.).

BUCK, N. G., LIGHT, D., RUTHERFORD, A., MILLER, M., RENNIE, T. W., PRATCHETT, D., CAPPER, B. S. and TRAIL, J. C. M. 1976. Environmental factors affecting beef cow reproductive performance in Botswana. *Anim. Prod.* **23:** 357–363.

CAPPER, B. S., PRATCHETT, D., RENNIE, T. W., LIGHT, D., RUTHERFORD, A., MILLER, M., BUCK, N. G. and TRAIL, J. C. M. 1977. Effects of rumen stimulatory licks on reproductive performance and live-weight gain of beef cattle in Botswana. *Anim. Prod.* **24:** 49–55.

CROCKETT, J. R., KOGER, M. and FRANKE, D. E. 1978a. Rotational crossbreeding of beef cattle: reproduction by generation. *J. Anim. Sci.* **46:** 1163–1169.

CROCKETT, J. R., KOGER, M. and FRANKE, D. E. 1978b. Rotational crossbreeding of beef cattle: preweaning traits by generation. *J. Anim. Sci.* **46:** 1170–1177.

CUNDIFF, L. V., GREGORY, K. E. and KOCH, R. M. 1974a. Effects of heterosis on reproduction in Hereford, Angus and Shorthorn cattle. *J. Anim. Sci.* **38:** 711–727.

CUNDIFF, L. V., GREGORY, K. E., SCHWULST, F. J. and KOCH, R. M. 1974b. Effects of heterosis on maternal performance and milk production in Hereford, Angus and Shorthorn cattle. *J. Anim. Sci.* **38:** 728–745.

DEESE, R. E. and KOGER, M. 1967. Heritability of fertility in Brahman and crossbred cattle. *J. Anim. Sci.* **26:** 984–987.

HARWIN, G. O., LAMB, R. D. and BISSCHOP, J. H. R. 1967. Some factors affecting reproductive performance in beef females. *Proc. S. Afr. Soc. Anim. Prod.* **6:** 171–177.

HOLNESS, D. H., HOPLEY, J. D. H. and HALE, D. H. 1978. The effects of plane of nutrition, live weight, temporary weaning and breed on the occurrence of oestrus in beef cows during the post-partum period. *Anim. Prod.* **26:** 47–54.

KOCH, R. M. 1972. The role of maternal effects in animal breeding: VI. Maternal effects in beef cattle. *J. Anim. Sci.* **35:** 1316–1323.

KOGER, M. 1973. Summary. In *Crossbreeding Beef Cattle*, Ser. 2 (ed. M. Koger, T. J. Cunha and A. C. Warnick), pp. 434–447. University of Florida Press, Gainesville, Fla.

MCDONALD, R. P. and TURNER, J. W. 1972. Estimation of maternal heterosis in preweaning traits of beef cattle. *J. Anim. Sci.* **35:** 1146–1154.

PRESTON, T. R. and WILLIS, M. B. 1974. *Intensive Beef Production.* 2nd ed. Pergamon Press, Oxford.

RAKHA, A. M., IGBOELI, G. and KING, J. L. 1971. Calving interval, gestation and postpartum periods of indigenous Central African cattle under a restricted system of breeding. *J. Anim. Sci.* **32:** 507–509.

RICHARDSON, F. D., OLIVER, J. and CLARKE, G. P. Y. 1975. Analyses of some factors which affect the productivity of beef cows and of their calves in a marginal rainfall area of Rhodesia. 1. Factors affecting calving rate. *Anim. Prod.* **21:** 41–49.

SMITH, G. M., FITZHUGH, H. A., CUNDIFF, L. V., CARTWRIGHT, T. C. and GREGORY, K. E. 1976. Heterosis for maturing patterns in Hereford, Angus and Shorthorn cattle. *J. Anim. Sci.* **43:** 380–388.

THORPE, W., CRUIKSHANK, D. K. R. and THOMPSON, R. 1979. The growth and carcass characteristics of crosses of Hereford and Friesian with Angoni, Barotse and Boran cattle in Zambia. *J. agric. Sci., Camb.* **93:** 423–430.

THORPE, W., CRUICKSHANK, D. K. R. and THOMPSON, R. 1980a. Genetic and environmental influences on beef cattle production in Zambia. 1. Factors affecting weaner production from Angoni, Barotse and Boran dams. *Anim. Prod.* **30:** 217–234.

THORPE, W., CRUICKSHANK, D. K. R. and THOMPSON, R. 1980b. Genetic and environmental influences on beef cattle production in Zambia. 2. Live weights for age of purebred and reciprocally crossbred progeny. *Anim. Prod.* **30:** 235–243.

THORPE, W., CRUICKSHANK, D. K. R. and THOMPSON, R. 1980c. Genetic and environmental influences on beef cattle production in Zambia. 3. Carcass characteristics of purebred and reciprocally crossbred castrated males. *Anim. Prod.* **30:** 245–252.

TONN, R. 1975. Selection criteria in extensive breeding of beef cattle with examples from Kenya. *Tierzüchter* **27:** 425–428.

TRAIL, J. C. M., BUCK, N. G., LIGHT, D., RENNIE, T. W., RUTHERFORD, A., MILLER, M., PRATCHETT, D. and CAPPER, B. S. 1977. Productivity of Africander, Tswana, Tuli and crossbred beef cattle in Botswana. *Anim. Prod.* **24:** 57–62.

TRAIL, J. C. M., SACKER, G. D. and FISHER, I. L. 1971. Crossbreeding beef cattle in Western Uganda. 3. Genetic analysis of body weight. *Anim. Prod.* **13:** 153–163.

WARD, H. K. 1968. Supplementation of beef cows grazing on veld. *Rhod. J. agric. Res.* **6:** 93–101.

(*Received* 27 *August* 1980—*Accepted* 21 *January* 1981)

36

The growth and carcass characteristics of crosses of Hereford and Friesian with Angoni, Barotse and Boran cattle in Zambia

by

W. Thorpe, D.K.R. Cruickshank and R. Thompson

J. agric. Sci., Camb. (1979), **93**, 423–430
Printed in Great Britain

The growth and carcass characteristics of crosses of Hereford and Friesian with Angoni, Barotse and Boran cattle in Zambia

By W. THORPE, D. K. R. CRUICKSHANK
Central Research Station, Mazabuka, Zambia
AND R. THOMPSON
A.R.C. Unit of Statistics, University of Edinburgh

(*Revised MS. received* 30 *March* 1979)

SUMMARY

Seven hundred and fifty purebred Angoni, Barotse and Boran cattle and their crosses sired by Friesian and Hereford bulls were compared for live-weight and carcass characters in Zambia. The cattle were born between September and December in three consecutive years and were reared on unimproved veld. In the first year all females and a random half of the males received supplementary feed in winter and the males were slaughtered at 2·5 years of age. The remaining males received the same total amount of winter supplementary feed prior to slaughter at 3·0 years of age. The males born in the second year were pen fed for a period prior to slaughter at 3·0 years of age. Males born in the third year were slaughtered at 3·5 years of age and received no supplementary feed.

Genotype differences were important at all live weights and there were year of birth × genotype interactions at the 0·5, 1·0 and 1·5 year weights and a genotype × feed treatment interaction at the 2·0 year weight. Boran purebreds were heavier than the Angoni and Barotse purebreds; at 3·0 years the differences were +42·3 kg (+12·9 %) and +25·2 kg (+7·7 %) respectively. Cross-breds had heavier live weights than their corresponding purebreds with Boran crosses consistently heavier than the Angoni and Barotse crosses. The Friesian and Hereford sired Boran cross-breds were +51·0 kg (+13·5 %) and +43·0 kg (+11·6 %) heavier respectively than the purebred Borans at 3·0 years of age.

There were large differences between the genotypes for all carcass characters with Borans having heavier carcasses than the Angonis and Barotses by +20·3 kg (+10·6 %) and +19·1 kg (+10·0 %) respectively. The average carcass weight advantages of Friesian and Hereford crosses with the Angoni, Barotse and Boran breeds over the corresponding purebreds were +40·3 kg (+19·0 %), +33·6 kg (+16·2 %) and +31·0 kg (+14·0 %) respectively. There were few important carcass differences between the corresponding Friesian and Hereford crosses.

The mortality rate during the experiment was 8·2 % with only small differences between genotypes.

The interactions found in the experiment indicated that breed comparisons should be conducted over a number of years and at various management levels. For growth and carcass production under the conditions of the experiment the use of exotic cross-breds is clearly advantageous.

INTRODUCTION

Genetic improvement can be expected to play a part in increasing the efficiency of cattle production even in the absence of concurrent improvements in management. Although these conditions often prevail in the developing countries of Africa, reliable estimates of breed performance and comparisons among breeds which would provide the basis for genetic improvement are seldom available. In Zambia for example, published estimates are limited to a study of the indigenous Angoni (Walker, 1964) and a comparison of the Angoni with exotic breeds for some reproductive charac-

0021-8596/79/2828-2450 $01.00 © 1979 Cambridge University Press

ters (Rakha, Hale & Igboeli, 1970; Rakha, Igboeli & King, 1971).

The present paper is concerned with the growth and carcass characteristics of crosses of the Hereford and Friesian breeds with Angoni, Barotse and Boran cattle. The study was carried out by the Department of Agriculture in Zambia to provide an evaluation of the performance of breeds available in Zambia for ranching conditions and as a step towards the improved utilization of the genetic resources in that country. Breeds available with Zambia are numerous. They include the indigenous Sanga breeds, the Barotse and the Tonga and their crosses, and the indigenous Zebu, the Angoni. The former constitute some 60 % and the latter about 15 % of the total cattle population. Most of the cattle of these breeds (about 95 %) are maintained under a traditional management system. Grazing is communal on unenclosed land and management practices are few and of a low standard. In the commercial sector, which comprises about 15 % of the national population, cattle are kept on fenced ranches. Management includes a restricted breeding season, the separation of breeding and young stock, disease control and the retention of standing hay for dry season grazing. Major breeds in the commercial sector are the Boran, a zebu type, and the Africander, a sanga type, originating from Ethiopia and South Africa respectively. Descriptions of these breeds have been given by Mason & Maule (1960). The Boran and Africander breeds are often cross-bred with the British beef breeds, the Hereford and Sussex. The small commercial dairy industry is based on the Friesian breed imported from Britain, Kenya and South Africa and there is a growing interest in the Friesian × indigenous female for milk production.

MATERIALS AND METHODS

The cattle compared were born in three years, 1969, 1970 and 1971. Live-weight records were collected at the Central Research Station, Mazabuka, Southern Province, Zambia which is situated at an altitude of 990 m on the edge of the peneplain of the Kafue River, latitude 15 ° 50′ S and longitude 27° 45′ E. The station has a mean annual temperature of 22·0 °C. Annual rainfall is seasonal, falling mainly between October and April. Annual and monthly rainfall totals varied widely over the period of the breed comparison, 1969–75 (Table 1).

In Zambia most grazing is natural veld and herbage growth is a reflexion of the seasonal rainfall distribution with plentiful, fresh, high quality grazing in the early months of the rains followed by rapid plant maturity and a sudden reduction in the veld quality. As a result dry season grazing is low quality, mature herbage. The fluctuating nutritional level gives rapid weight gains during the wet season and maintenance or submaintenance nutrition in the dry season (Smith, 1959).

The breeding season was restricted to a 3 month period from late December to early March and the dams of the experimental cattle grazed large, well-watered upland Hyparrhenia veld paddocks except during the 1970 calving season when they grazed adjoining river flats. Male calves were castrated in April and all calves were weaned at about 8 months of age within 3 days in early June. After a short period of pen feeding the weaners grazed the upland Hyparrhenia veld. Health precautions included weekly spraying with acaricides. Rock salt was usually available.

The numbers of live-weight records, classified by their year of birth, genotype and sex, are given in Table 2. The purebred Angoni, Barotse and Boran calves were sired by 12, 13 and 14 bulls respectively and the Friesian and Hereford cross-bred calves by five and eight bulls respectively.

Data on 750 records were analysed for birth and 0·5 year weights, 594 records for 1·0 and 1·5 year weights and 423 records for 2·0 and 2·5 year weights. The reduced number of records at the

Table 1. *Monthly rainfall totals (mm) at the Central Research Station, Mazabuka, 1969–75*

	1969–70	1970–71	1971–72	1972–73	1973–74	1974–75	Mean
August	0	0	0	0	0	0	0
September	0	0	2	15	0	0	3
October	36	16	2	16	71	1	24
November	96	76	163	18	111	97	93
December	380	106	207	95	233	300	220
January	164	222	226	134	235	138	186
February	93	172	121	181	249	164	164
March	18	20	80	4	41	59	37
April	12	2	18	18	18	0	11
May	0	0	0	0	0	0	0
June	0	0	0	0	0	0	0
July	0	0	0	0	0	0	0
Total	799	614	819	481	958	759	739

Table 2. *Number of live-weight records classified by year of birth, genotype and sex*

	Year of birth			
	1969	1970	1971	Total
Sex				
Female	90	167	146	403
Male castrate	124	69	154	347
Genotype				
Angoni	39	59	56	154
Barotse	34	43	56	133
Boran	34	45	49	128
Friesian × Angoni	28	17	20	65
Friesian × Barotse	26	8	11	45
Friesian × Boran	14	14	18	46
Hereford × Angoni	20	19	38	77
Hereford × Barotse	4	16	22	42
Hereford × Boran	15	15	30	60
Total	214	236	300	750

Table 3. *Number of male castrate carcass records classified by year of birth, genotype and slaughter age*

	Year of birth				
	1969*				
Slaughter age/feeding treatment ...	2·5	3·0	1970 3·0†	1971 3·5‡	Total
Genotype					
Angoni	12	12	5	30	59
Barotse	11	10	11	19	51
Boran	10	10	6	19	45
Friesian × Angoni	7	8	6	8	29
Friesian × Barotse	5	7	4	6	22
Friesian × Boran	4	3	9	10	26
Hereford × Angoni	4	4	10	16	34
Hereford × Barotse	2	0	9	14	25
Hereford × Boran	5	4	4	11	24
Total	60	58	64	133	315

* Veld grazing plus dry season supplementary feed.
† Pen feeding prior to slaughter.
‡ Veld grazing only.

later ages was mainly a consequence of females not being available for the comparison after 0·5 year for those born in 1971 and after 1·5 years for those born in 1970. Mortalities were not an important cause of the reduced number of records. The mortality rates during the experiment are presented in the results section. No females were slaughtered but a random half of the 1969 born males were slaughtered at 2·5 years, leaving a total of 265 male records for the 3·0 year weight analysis. The post-weaning live weights were recorded on the day approximating to the date on which the mean age of the group corresponded to 1·0, 1·5, 2·0 years etc. The live weights at 1·0, 2·0 and 3·0 years generally corresponded to the minimum live weight recorded at the end of the dry season and the half yearly weights, generally recorded in late May, were usually the maximum live weights attained after the rainy season.

Different management practices were applied to the three year-of-birth groups and animals in the slaughter age/feeding treatment groups were slaughtered at a common time rather than at a constant degree of finish or at a fixed live weight. For the 1969 born group all females and about half of the males, which were randomly chosen, were given supplementary feed for 145 days while grazing veld during their second dry season; that is, just after 2·5 years of age. The supplement was a mixture containing 17 % cottonseedmeal and 83 % snapcorn or maize bran fed on a live-weight scale. The males were fed again for 40 days prior to slaughter at 2·5 years of age. The remaining random sample of 1969 born males was supplementary fed the same total amount of feed as the earlier group during 90 days before slaughter at 3·0 years of age. The 1970 born males were pen fed for 110 days prior to slaughter at 3·0 years of age while the 1971 born males received no supplementary feed and were slaughtered at 3·5 years of age. The numbers of males slaughtered for each genotype, year of birth and slaughter age/feeding treatment group are given in Table 3.

Final live weights were recorded without starving the animals which were slaughtered usually the day after arrival at the Lusaka abattoir. Wet hide weight was recorded at the time of slaughter and carcasses were dressed and split in the normal manner. The methods used for quartering, separation of the tenth rib joint and the hump and the specific gravity determination of the former have been described by Ledger, Gilliver & Robb (1973). Agricultural Research Council standards were used to score the chilled carcasses for carcass finish and leg shape and linear measurements including the average of the fat depths, C, D and E, were made (Agricultural Research Council, 1965).

Genotype effects were estimated by least squares analysis using appropriate models. For the characters birth weight to 1·5 year weight, the effects fitted were the animal's year of birth (1969, 1970, 1971), period of birth (< 294, 294–314, $\geqslant 315$ days), sex (female, male), genotype (Angoni, Barotse, Boran, Friesian × Angoni, Friesian × Barotse, Friesian × Boran, Hereford × Angoni, Hereford × Barotse, Hereford × Boran), dam age (3, 4, 5 and 6, 7 and 8, 9, $\geqslant 10$ years), dam previous parous state (heifer, calved previous year, barren previous year), the two factor interactions of major effects and a regression for date of birth. As only males were represented at the 2·0 and 2·5 year weights in the 1970 and 1971 year of birth groups the sex effect was confounded with feed treatment

differences. At the 3·0 year weight feed treatment and year of birth effects were confounded. For the carcass characters of castrated males, a compound effect representing the confounded year of birth, feed treatment and slaughter age effects was fitted along with the effects period of birth, genotype and the interaction of genotype with block and a regression for date of birth.

RESULTS

Live weights. Genotype was generally highly significant for most live weights and year of birth × genotype interactions were highly significant for the 0·5, 1·0 and 1·5 year weights. Differences between the purebred mean live weights were consistent in the 3 year of birth groups at all weights but the differences between the cross-breds and the purebreds varied up to 2·0 years of age. Some rank changes occurred between the cross-breds but the cross-bred live weights were generally superior to those of their corresponding purebreds. Table 4 gives the least square estimates of comparative breed performances adjusted for all effects.

Boran purebreds had heavier live weights than either the Angoni or Barotse purebreds. At 3·0 years of age the advantage was +42·3 kg (+12·9 %) and +25·2 kg (+7·7 %) respectively. The superiority was generally highly significant, as was the superiority of the Barotse over the Angoni. The cross-breds had heavier live weights than their corresponding purebreds. For example, Friesian and Hereford sired Boran cross-breds were +51·0 kg (+13·5 %) and +43·0 kg (+11·6 %) heavier respectively than the purebred Borans at 3·0 years of age. The differences were generally highly significant. Boran crosses were consistently significantly heavier than the Barotse and Angoni crosses which were rarely significantly different. The Friesian sired crosses had heavier live weights than the comparable Hereford crosses but the differences were only statistically significant for the progeny of Barotse dams.

At the 2·0 year weight the 1969 born group gave a significant genotype × feed treatment interaction. The subclass means (not tabulated) showed that the live weight difference between the *Bos taurus* cross-breds and the local purebreds was increased when supplementary feed was given in the dry season. The differential effect was greater with the Hereford sired steers which, when fed, were on average 42 kg heavier than the fed purebred steers. The unfed Hereford crosses were only 7 kg heavier than the unfed purebreds. There was no interaction at the 2·5 year weight which was recorded at the end of the period of abundant rainy season grazing. Compensatory growth could be expected during the period between the 2·0 year and 2·5 year weights.

Carcass characters. There were highly significant environmental (year of birth/feed treatment/slaughter age) and genotype effects for all carcass characters excepting a compound effect on carcass weight. There were no compound × genotype interactions. The least squares genotype means are presented in Table 5.

Among the purebreds the superior live weight of the Boran combined with a relatively high killing out percentage to give a carcass significantly heavier than the carcasses of the Angoni and Barotse purebreds. The advantages were +20·3 kg (+10·6 %) and +19·1 kg (+10·0 %) respectively. Angoni carcasess had a higher killing out percentage which removed the significant advantage shown

Table 4. *Least squares mean live weights (kg) of Angoni, Barotse and Boran purebreds and their Friesian and Hereford sired crosses from birth to 3·0 years of age*

	Age (years)						
	Birth	0·5	1·0	1·5	2·0	2·5	3·0
Genotype							
Angoni	21·1	121·5	125·9	191·5	189·5	280·9	285·5
Barotse	24·0	132·6	141·7	208·7	201·1	298·2	302·6
Boran	25·7	145·8	149·3	222·2	227·7	324·5	327·8
Friesian × Angoni	24·7	141·6	153·1	237·8	235·0	338·3	362·6
Friesian × Barotse	28·8	151·2	166·6	247·2	244·9	345·5	358·7
Friesian × Boran	27·4	158·0	172·7	257·0	262·6	360·9	378·8
Hereford × Angoni	23·5	140·7	149·4	234·7	233·5	337·4	351·8
Hereford × Barotse	25·7	139·7	149·8	232·3	223·6	332·2	348·6
Hereford × Boran	26·0	156·5	162·8	252·6	246·0	350·3	370·8
Approximate S.E.s of differences between means							
Purebred – purebred	0·40	2·12	2·56	3·19	4·45	5·10	6·39
Cross-bred – purebred	0·55	2·94	3·39	4·23	5·85	6·71	8·22
Cross-bred – cross-bred	0·65	3·46	3·94	4·92	6·49	7·45	8·81

Table 5. *Least squares carcass character means of Angoni, Barotse and Boran breeds and their Friesian and Hereford sired crosses*

Genotype ...	Angoni	Barotse	Boran	Friesian × Angoni	Friesian × Barotse	Friesian × Boran	Hereford × Angoni	Hereford × Barotse	Hereford × Boran	Approximate S.E.s of differences between means: Pure –pure	Cross –pure	Cross –cross
Slaughter wt. (kg)	322·8	345·6	367·0	408·9	412·3	420·7	397·3	388·6	420·9	6·27	7·78	8·87
Carcass wt. (kg)	171·0	172·2	191·3	211·4	212·6	220·9	211·1	198·9	223·6	3·90	4·83	5·51
Killing out (%)	53·0	50·0	52·2	51·7	51·6	52·5	53·1	51·3	53·1	0·45	0·56	0·64
Hide (% of live weight)	7·9	8·2	9·4	7·0	7·4	7·5	7·8	7·8	8·2	0·11	0·18	0·21
Forequarter (% of carcass)	46·7	46·4	46·8	46·8	46·2	46·7	46·8	47·2	46·7	0·36	0·45	0·51
Hump (% of carcass)	0·94	0·45	0·78	0·44	0·31	0·39	0·46	0·43	0·44	0·033	0·041	0·046
Forearm length (cm)	35·2	37·4	37·2	37·7	38·4	38·3	36·6	37·4	36·5	0·30	0·38	0·43
Carcass width (cm)	43·3	45·0	44·9	46·9	46·4	46·9	47·0	45·8	47·1	0·31	0·38	0·44
Carcass length (cm)	109·7	114·1	115·2	120·5	122·8	122·7	117·6	117·3	119·6	0·73	0·90	1·03
Leg shape (score)†	2·43	2·14	2·45	2·42	2·16	2·43	2·96	2·60	3·18	0·12	0·15	0·17
Carcass finish (score)‡	2·45	2·15	2·69	2·05	1·80	2·19	2·71	2·31	3·09	0·12	0·15	0·17
Average fat depth (mm)	9·6	6·2	8·0	6·5	5·3	7·1	7·6	6·3	8·4	0·46	0·57	0·65
'Eye muscle' $A \times B$ (mm²)	59·4	55·8	58·3	63·5	59·6	66·4	58·6	58·6	62·8	1·63	2·02	2·30
Specific gravity 10th rib joint	1·055	1·078	1·060	1·071	1·080	1·070	1·062	1·069	1·057	0·006	0·008	0·009
Fat (%)*	21·8	15·0	20·3	17·1	14·5	17·4	19·7	17·7	21·2	—	—	—
Lean (%)*	61·2	64·7	62·0	63·7	65·0	63·6	62·4	63·4	61·5	—	—	—
Bone (%)*	17·0	20·3	17·6	19·2	20·6	19·1	17·8	18·9	17·2	—	—	—

* Derived from the results in Ledger, Gilliver & Robb (1973).

† Leg shape score: 1 (poor) to 5 (good).

‡ Carcass finish score: 1 (poor) to 5 (good).

Table 6. *Least squares carcass character means of male castrates slaughtered at two ages*

Slaughter age (year)	Slaughter wt. (kg)	Carcass wt. (kg)	Carcass length (cm)	'Eye muscle' $A \times B$ (mm²)	Carcass finish (score)	Fat depth (mm)
2·5	383·8	196·8	117·1	63·7	2·40	7·1
3·0	389·0	201·3	117·2	62·9	2·45	8·1
S.E. of difference	5·94	3·69	0·69	1·54	0·11	0·44

at slaughter weight by the Barotse. While all the purebreds differed highly significantly for hump percentage, there were no differences in forequarter proportion. The small Angoni gave significantly shorter, narrower carcasses with shorter forearms than the Barotse and the Boran which did not differ significantly. The Barotse had significantly lower (poorer) leg and carcass finish scores, subjective measures used in carcass grading, and less fat depth than the two zebu breeds and a significantly smaller 'eye muscle' area than the Angoni. The Angoni and Boran purebreds did not differ significantly for the 10th rib joint specific gravity and calculated tissue percentages showed the Barotse carcasses to have a higher proportion of lean and bone and a lower proportion of fat than the zebu breeds.

All cross-bred slaughter weights were highly significantly heavier than the weights of the corresponding purebreds, differences which were maintained at carcass weight and repeated in the longer linear measurements of the cross-breds. The average carcass weight advantages of the Friesian and Hereford crosses with the Angoni, Barotse and Boran breeds over the corresponding purebreds were +40·3 kg (+19·0 %), +33·6 kg (+16·2 %) and +31·0 kg (+14·0 %) respectively. Cross-bred hump percentages were generally significantly smaller than, and forequarter percentages similar to, those found in the purebreds. Hereford but not Friesian crosses had better leg and carcass finish scores than the purebreds. The Barotse and Boran crosses did not differ significantly from their corresponding purebreds for fat depth. Only the Friesian × Angoni 10th rib joint specific gravity differed significantly from the corresponding purebred result. Derived tissue percentages showed the crosses to have generally higher proportions of lean and bone and less fat when slaughtered at a common time.

There were few important carcass differences between the corresponding Friesian and Hereford crosses except that Friesian × Barotse steers gave significantly heavier carcasses than Hereford × Barotse steers and Friesian sired progeny tended to be less well finished.

The influence of the compound effect was not statistically significant for carcass weight due to its contrasting effects on slaughter weight and killing out percentage. Within the 1969 born group carcass weights and linear measurements were similar but fatness differed when steers were fed to be slaughtered at 2·5 and 3·0 years of age (Table 6).

Mortality rate. The mortality rate during the experiment was 8·2 % with only small differences between genotypes. Losses were approximately the same pre- and post-weaning for the *Bos taurus* sired progeny, 3·2 and 4·6 % respectively, but the pre-weaning mortality of the purebred progeny, 6·1 %, was about double their post-weaning mortality, 2·6 %. The mortalities arose from a variety of causes with no single cause predominant.

Reproductive rate. Comparable reproductive performances of the Friesian sired F_1 females and purebred females were not available as all Friesian sired females were used in a cross-bred dairy experiment. The management and nutrition in the experiment were similar to those provided for the ranch herds and the calving percentages achieved were competitive with those of the purebred ranch dams. The 1970 and 1971 born Hereford sired female groups were not retained for breeding. However the small number of females born in 1969 were bred subsequently as contemporary ranch cows along with the local purebreds and their cross-breds. The calving percentages and calf weaning weights of the Hereford F_1 cross-breds were superior to those of the other genotypes. It was noted that the *Bos taurus* sired F_1 females had earlier puberty than the local purebred females. These reproductive results will be reported separately.

DISCUSSION

It has long been assumed in southern Africa that under conditions of reasonable management African cattle breeds are not competitive with their crosses with European breeds. The validity of that assumption for growth and carcass characters is confirmed by the present results. In addition the results indicate the improvements possible by replacing the Zambian breeds, Angoni and Barotse, by the East African breed Boran which gave on average 10 % more carcass weight than the Zambian breeds. Further improvements

resulted when the three African breeds were mated with Friesian and Hereford sires. The cross-bred progeny were superior to the Angoni, Barotse and Boran purebreds giving, on average, +19, +16 and +14% more carcass weight respectively. For growth and carcass production under the conditions of the experiment the use of exotic breeds for cross-breeding is clearly advantageous. In Kenya similar carcass weight advantages for cross-bred progeny have been reported when Boran dams were crossed with Herefords and Friesians (Tonn, 1974). Smaller increases in carcass weight resulted when Angus and Red Poll bulls were crossed with Boran, Ankole and Zebu dams in Uganda (Trail, Sacker & Marples, 1971). Live-weight advantages for Simmental crosses with local breeds at weaning in Kenya (C. R. Kamau, personal communication) and at 18 months of age in Botswana (Trail *et al.* 1977) suggest that carcass weight advantages will occur with crosses sired by the large European beef breeds. In the present experiment the greatest improvement could be achieved by utilizing the *Bos taurus* × Boran steers which gave over 22% more carcass weight than the indigenous purebred steers.

Kebede *et al.* (1977) have reported a breed comparison in Ethiopia in which Friesian, Simmental and Jersey sires were mated to three zebu breeds including the Boran. Sire breed × dam breed interactions for progeny live weights indicated the possibility of selecting specific breed combinations to achieve performances above those expected from the breed means. In the present comparison there were no sire breed × dam breed interactions. However when selecting between sire breeds it is necessary to consider the possibility of specific combining abilities if breed differences are to be optimized. Comparisons between the crosses of the alternative sire breeds with all available dam breeds are required.

Trail (1969) has reported year × genotype interactions for live-weight. In the present experiment there were year × genotype and genotype × feed treatment interactions. Interactions in the slaughter generation will be practically important only if they also occur for carcass characters and the confounding of year of birth and management treatments in the 1970 and 1971 born groups precludes any definite conclusions being drawn about their importance. Nevertheless the observed genotype × year and genotype × feed treatment interactions for live weights suggest that breed comparisons should be conducted over a number of years and with various management systems. Comparisons made with alternative management systems will indicate any advantages of specific genotype × management system combinations.

While large advantages for carcass production have been shown to accrue from replacing local breeds by *Bos taurus* × Boran steers, final recommendations on inter-breed selection and cross-breeding must consider all phases of the production cycle. Turner (1967) has emphasized the importance of 'fitness' characteristics in the evaluation and utilization of indigenous breeds and Rennie *et al.* (1977) have shown that low reproductive performance and maternal ability are the main factors that depress productivity of cattle in Botswana. Comparative mortality rates, relative reproductive efficiencies and lifetime productivities of purebred and cross-bred progeny must be considered. The Zambian results so far available suggest that the reproductive rates of Friesian and Hereford sired F_1 females are competitive with the performance of the local purebreds and their *inter se* crosses. Their earlier puberty is in line with the results of Trail (1969) and is to be expected from the results of Rakha *et al.* (1970). Mortality rates of the exotic cross-breds have also been shown to be competitive with the local purebreds. However further comparisons are required to substantiate these results and to give lifetime cow productivities.

If these preliminary fitness results are substantiated for the exotic cross-breds, when they are taken together with the large advantage in carcass weights there is much to recommend the utilization of exotic cross-breds for ranching in Zambia. Theoretical and practical considerations suggest that a criss-cross breeding system may be appropriate (Koger, 1973).

The authors wish to thank the Director of Agriculture for the provision of experimental facilities. They acknowledge the technical assistance of C. D. Slater, P. J. Hatakwati and all Animal Husbandry Research staff at Mazabuka. The help of Veterinary Department staff was much appreciated. The UK Ministry of Overseas Development provided financial support for the analyses through research scheme R3390.

REFERENCES

AGRICULTURAL RESEARCH COUNCIL (1965). *Recommended Procedures for Use in the Measurement of Beef Cattle and Carcasses.* London: Agricultural Research Council.

KEBEDE, B., GALAL, E. S. E., WOLDE, A. G., O'DONOVAN, P. B., GEBRELUL, S., WOLDEGEBRIEL, A. & MARIAM, H. G. (1977). Study on growth rate of F_1 calves. *Animal Production Bulletin No. 1. Institute of Agricultural Research, Addis Ababa.*

KOGER, M. (1973). Alternative procedures in crossbreeding. In *Crossbreeding Beef Cattle*, series 2 (ed. M. Koger, T. J. Cunha and A. C. Warnick), pp. 448–453. Gainesville: University of Florida Press.

LEDGER, H. P., GILLIVER, B. & ROBB, J. M. (1973). An examination of sample joint and specific gravity techniques for assessing the carcass composition of steers slaughtered in commercial abattoirs. *Journal of Agricultural Science, Cambridge* **80**, 381–392.

MASON, I. L. & MAULE, J. P. (1960). *The Indigenous Livestock of Eastern and Southern Africa.* Commonwealth Agricultural Bureau.

RAKHA, A. M., HALE, D. & IGBOELI, G. (1970). The age of puberty in local breeds of cattle in Central Africa. *Journal of Reproduction and Fertility* **22**, 369–370.

RAKHA, A. M., IGBOELI, G. & KING, J. L. (1971). Calving interval, gestation and postpartum periods of indigenous Central African cattle under a restricted system of breeding. *Journal of Animal Science* **32**, 507–509.

RENNIE, T., LIGHT, D., RUTHERFORD, A., MILLER, M., FISHER, I., PRATCHETT, D., CAPPER, B., BUCK, N. & TRAIL, J. (1977). Beef cattle productivity under traditional and improved management in Botswana. *Tropical Animal Health and Production* **9**, 1–6.

SMITH, C. A. (1959). Studies on the Northern Rhodesian Hyparrhenia veld. Part 1. The grazing behaviour of indigenous cattle grazed at light and heavy stocking rates. *Journal of Agricultural Science, Cambridge* **52**, 369–375.

TONN, R. (1974). The suitability of the Boran and its crosses with European breeds for range beef production in the tropics, investigated on ranches in the Rift Valley of Kenya. Thesis, Georg-August-Universitat, Gottingen.

TRAIL, J. C. M. (1969). A study of performance traits of crossbred cattle in a new beef cattle ranching area of Uganda. Ph.D. Thesis, University of Edinburgh.

TRAIL, J. C. M., BUCK, N. G., LIGHT, D., RENNIE, T. W., RUTHERFORD, A., MILLER, M., PRATCHETT, D. & CAPPER, B. S. (1977). Productivity of Africander, Tswana, Tuli and crossbred beef cattle in Botswana *Animal Production* **24**, 57–62.

TRAIL, J. C. M., SACKER, G. D. & MARPLES, H. J. S. (1971). Crossbreeding beef cattle in Western Uganda. 5. Growth and carcass evaluation of castrated males. *Animal Production* **13**, 171–180.

TURNER, H. N. (1967). Genetic aspects of characteristics to be considered in the evaluation and utilization of indigenous livestock. *World Review of Animal Production* **3**, 17–26.

WALKER, C. A. (1964). The growth and development of the beef qualities of Angoni cattle (East African zebu). *Journal of Agricultural Science, Cambridge* **63**, 135–172.

37

Genetic parameters of beef characteristics and the genetic relationship between meat and milk production in British Friesian cattle

by

I.L. Mason, V.E. Vial and R. Thompson

Anim. Prod. 1972, **14**: 135–148

GENETIC PARAMETERS OF BEEF CHARACTERS AND THE GENETIC RELATIONSHIP BETWEEN MEAT AND MILK PRODUCTION IN BRITISH FRIESIAN CATTLE

I. L. MASON†

ARC Unit of Animal Genetics, Institute of Animal Genetics, Edinburgh EH9 3JN

V. E. VIAL

The Agricultural Institute, Dunsinea, Castleknock, Co. Dublin, Ireland

AND

R. THOMPSON

ARC Unit of Statistics, Edinburgh EH8 9LN

SUMMARY

A beef progeny test was organized for 66 British Friesian bulls standing at AI centres in England. Eighty-one progeny groups averaging 8·5 steers each were reared on the farms of various co-operating organizations and their growth and carcass characters were recorded. Heritability of various measurements of gain varied from 0 to 0·34, of carcass conformation measures from 0·24 to 0·50, of muscle and fat percentage in the rib joint was 0·58 and 0·46 respectively, of other measures of fatness varied from 0·04 to 0·18, of weight of feet was 0·4 and of weight of hide 0·5.

Phenotypic correlations among beef characters gave no indication that carcass composition could be predicted from the external carcass measurements or subjective scores used in this study. Genetic correlations suggested that selection for weight for age would produce a fatter and more compact carcass. Genetic correlations were high and negative between milk yield (contemporary comparison) and weight for age, carcass weight/length, and blockiness of leg, and high and positive between milk yield, and carcass depth/length and percentage bone in the rib joint. Their standard errors were also high.

INTRODUCTION

A PREVIOUS paper (Mason, 1964) posed and discussed the question: what is likely to be the correlated response in growth rate and carcass quality to selection for milk yield? Analysis of data from commercial herds of Red Poll and Dairy Shorthorn cattle gave low and statistically insignificant genetic correlations between growth rate and dressing percentage, on the one hand, and milk yield and fat percentage on the other. In an attempt to obtain more reliable figures from a more numerous breed, a beef progeny test for British Friesian bulls at artificial insemination (AI) centres was organized.

† Present address: Animal Production and Health Division, FAO, Via delle Terme di Caracalla, 00100 Rome, Italy.

MATERIAL AND METHODS

Animals

The aim was to use 10 male progeny by each of a number of bulls standing at the same AI centre in England at the same time and to repeat the test at several centres and over a number of years. The progeny were all normal, full-term, single calves out of Friesian or Friesian-type cows. They were born during the years 1955–61 and, with a few exceptions, between mid-August and the end of November, mostly in September and October. They were collected within a week of birth and reared together on the farm of one of the cooperating organizations (see Acknowledgements). All animals were castrated (steers). In all, 66 bulls were tested; eight were tested twice, two three times and one four times so that 81 progeny groups were collected. There were seven rearing centres which cooperated for one or more years making a total of 28 centre-years. At 11 centre-years sires were tested two at a time, at 11 in groups of three, at five in groups of four, and one centre in one year accommodated six progeny groups. The progeny groups varied in size from 4 to 13 with a mean of 8·5.

At any centre in any one year all steers were treated alike, but at different centres rearing conditions were not necessarily the same. Four progeny groups (two centre-years) were kept inside and fed fairly intensively for slaughter at 15 to 19 months of age. Forty progeny groups were kept on semi-intensive systems being indoors during the first winter, on grass during the summer and fed in yards during the second winter to be slaughtered in the spring (usually March to May) at an age of 17 to 20 months. Ten progeny groups (three centre-years) were 'stored' (i.e. kept at a maintenance or sub-maintenance level on roughages) during the second winter, finished in the following summer on grass and slaughtered between June and October at the age of 20 to 26 months. The remaining 27 progeny groups (eight centre-years) were managed extensively and slaughtered at 25 to 30 months of age either off grass in late autumn and early winter or off silage in yards between January and March.

Steers were weighed (and usually also measured) on arrival, at 6 to 8 months of age and every 6 months thereafter; the exact time was regulated by the time of change of management in spring and autumn. These weights and measurements were made by the local staff. Animals were slaughtered either when they reached a predetermined weight (7 centre-years), when they reached a predetermined age (11 centre-years) or when the rearing centre considered them fit for slaughter. The constant weight subset included some centre-years for which the criterion was fitness for slaughter but in which the variance in weight was so low that it justified inclusion in the constant weight category. Usually the steers from any one centre went to slaughter in three or four batches over a period of 2 to 3 months but strict attention to age or weight entailed weekly slaughter and for constant weight the slaughter period sometimes extended to nearly 5 months.

Slaughter

At slaughter the caul (omental) fat was weighed as well as the wet hide and the four feet. The feet were removed at the knee joint between carpals and metacarpal and at the hock between tarsals and metatarsal. The carcass was weighed hot and the cold weight estimated by subtracting 2%. On the

day after slaughter linear measurements were taken on the hanging carcass and the measurements on the two sides were averaged. The amount of external fat cover was visually estimated and scored on a 1–9 scale; whenever possible it was scored by two observers and the scores averaged. The plumpness of the hind leg was similarly scored. These measurements and scores were made by V. E. V. or I. L. M. or K. W. Leach and in accordance with the recommendations of the Agricultural Research Council (1965). Carcasses were graded by the Livestock Inspectorate of the Ministry of Agriculture; their A, B and C grades were each subdivided into three and coded as C− = 0, C = 1, C+ = 2, B− = 3, B = 4, B+ = 5, A− = 6, A = 7, and A+ = 8; the select grade was coded S = 9.

After quartering between the 10th and 11th ribs the 10th rib joint was removed from the left side. The outline of the eye muscle (*m. longissimus dorsi*) was traced and its cross-sectional area measured from the tracing with a compensating planimeter. The thickness of subcutaneous fat was measured at three positions over the long axis of the eye muscle and the three measurements averaged. The whole rib joint from vertebral column to sternum (costal cartilage) was put in a polythene bag in a deep freeze and later dissected into separable fatty tissue ('fat'), muscular tissue ('muscle'), bone, and remainder (connective tissue, periosteum and cartilage).

The 10th rib was used as a sample joint since it could be conveniently removed from a carcass quartered between the 10th and 11th ribs with minimum reduction in the commercial value of the forequarter. Previous workers (Mason, 1951; Ledger and Hutchison, 1962) had shown that a complete single rib was as accurate a predictor of carcass composition as the prime ends of three ribs used in American work. (Mason calculated correlations, within breeds, between the composition of the complete 13th rib joint and that of the half-carcass of 0·92 for fat, 0·88 for lean and 0·55 for bone. Ledger and Hutchison, using the 10th rib of the Boran, obtained corresponding correlations of 0·96, 0·87 and 0·84. They showed that the correlations increased in size as the amount of the rib joint included in the dissection was increased.)

For 40 animals in the present material the rib was dissected in two halves—the plate end and the prime cut. The correlations between the composition of the two halves were: 0·81 for fat, 0·64 for lean and 0·35 for bone. In view of the inadequate size of these correlations it was decided to use the complete rib for dissection and not only the prime cut.

Owing to the difficulties of working on commercial farms and in commercial slaughterhouses not all measurements and observations were made on all steers and carcasses. A few missing values were estimated and inserted.

Characters

The weights and measurements were used to form the following characters used in the analysis:

Gain to 6 mo: daily live-weight gain from 7 days to 6–8 months of age.

Gain 6–12 mo: daily live-weight gain from 6–8 months to 12–14 months of age.

Gain 12–18 mo: daily live-weight gain from 12–14 months to 18 months of age.

Chest girth: body circumference immediately behind the shoulder blade, at 18 months of age.

Withers height: vertical height to the highest point of the withers between the shoulders, at 18 months of age.

Chest depth: vertical distance (with calipers) from point of withers to sternum, at 18 months of age.

Slaughter weight for age: starved live weight in kg immediately before slaughter, divided by age in days.

Carcass weight for age: weight of cold carcass in kg divided by age at slaughter in days.

Dressing %: cold carcass weight divided by live weight before slaughter.

Fleshing index: cold carcass weight in kg divided by length of side (in m) from anterior edge of H-bone (os pubis) to anterior edge of first rib.

Ranginess: depth of carcass from sternal cartilage to 7th vertebral spine cartilage divided by carcass length.

Blockiness of leg: length of hind leg from ridge on distal end of tibia to edge of cut fat in crutch divided by length from same ridge to anterior edge of H-bone.

The other characters have already been explained or are self-explanatory.

Statistical procedures

The within bull variance (σ^2_w) and between progeny groups variance (σ^2_s) were calculated from the analysis of variance by equating the within bull progeny groups and the between bull within centre-year sums of squares to their expectations. The heritability was estimated as $4\sigma^2_s/(\sigma^2_w+\sigma^2_s)$ and the standard error found using formulae developed by Searle (1961). Genetic and phenotypic correlations were calculated from the corresponding sums of cross-products. Approximate standard errors for genetic correlations were calculated according to the method of Tallis (1959).

The analyses on measurements and weight up to 18 months are based on different subsets of the data. For any particular trait only centre-years that had complete information on that trait were included in the analysis. Records were available from all slaughtered steers on weight/age, grade and fleshing index and analysis on these traits included all these 690 animals. For all the other variates related to the slaughtered animals the 541 animals that had information on all 18 traits were used.

The milk characteristics of the bulls were estimated using the first lactation of their daughters—heifer contemporary comparison for yield (*see* Robertson, Stewart and Ashton, 1956) and mean butterfat test. Since the measures of the bulls' milk and beef traits are from different animals in different environments the within-bull covariances, between milk and beef characteristics, are assumed to be zero. The covariances between progeny groups are then simply found by equating the between-bull within centre-year sums of cross-products to their expectation.

The genetic variance of milk yields for a particular sample of n bulls was found using the formula

$$\frac{1}{(n-1)_i}\sum_{=1}^{n}(m_i-\bar{m})^2-\frac{1}{n}\left(\sum_{i=1}^{n}\frac{1}{w_i}\right)\sigma^2_{wm}$$

with

$$\bar{m}=\left(\sum_{i=1}^{n}m_i\right)/n.$$

[Since $\sum_{i=1}^{n}(m_i-\bar{m})^2$ has expectation

$$n\sigma^2_{sm}+\left(\sum_{i=1}^{n}\frac{1}{w_i}\right)\sigma^2_{wm}-\frac{n\sigma^2_{sm}}{n}-\frac{1}{n}\left(\sum_{i=1}^{n}\frac{1}{w_i}\right)\sigma^2_{wm}$$

$$=(n-1)\sigma^2_{sm}-\left(\frac{n-1}{n}\right)\left(\sum_{i=1}^{n}\frac{1}{w_i}\right)\sigma^2_{wm}]$$

m_i and w_i are respectively the milk contemporary comparison and the effective number of daughters of the ith bull; σ^2_{wm} and σ^2_{sm} are the variances of milk yield within and between bulls respectively. A similar formula holds

TABLE 1

Variance and heritability of characters studied

Character	Mean	Variance within progeny groups (σ^2_w)	Variance between progeny groups (σ^2_s)	Heritability (h^2)	SE (h^2)
Live animal					
Gain to 6 mo (kg/day)	0·67	$98{\cdot}2\times10^{-4}$	$5{\cdot}3\times10^{-4}$	0·20	0·12a
Gain 6–12 mo (kg/day)	0·67	$99{\cdot}0\times10^{-4}$	$7{\cdot}7\times10^{-4}$	0·29	0·27b
Gain 12–18 mo (kg/day)	0·73	$157{\cdot}0\times10^{-4}$	$-0{\cdot}0\times10^{-4}$	−0·00	0·12c
Chest girth at 18 mo (cm)	173	36·1	−0·8	−0·09	0·13d
Withers height 18 mo (cm)	124	14·9	0·5	0·13	0·10d
Chest depth at 18 mo (cm)	67·6	8·1	0·1	0·05	0·12d
Slaughter wt/age (kg/day)	0·72	$23{\cdot}4\times10^{-4}$	$2{\cdot}2\times10^{-4}$	0·34	0·16
Carcass					
Weight/age (kg/day)	0·41	$9{\cdot}58\times10^{-4}$	$0{\cdot}46\times10^{-4}$	0·18	0·12e
Dressing %	57·4	2·48	0·38	0·53	0·19
Grade (score)	4·9	2·52	0·14	0·21	0·12e
Fleshing index (kg/m)	212	142	3·0	0·08	0·10e
Ranginess (depth as % length)	48·3	1·46	0·21	0·50	0·19
Blockiness (thigh as % leg)	34·2	2·27	0·28	0·45	0·18
Plumpness of leg (score)	4·7	1·24	0·08	0·24	0·15
Eye-muscle area (cm^2)	63·9	66·4	6·6	0·36	0·16
% muscle in rib joint	50·6	12·7	2·2	0·58	0·20
% bone in rib joint	16·7	2·65	0·05	0·08	0·11
Muscle/bone ratio in rib	3·09	0·149	0·009	0·22	0·15
% fat in rib joint	26·5	16·5	2·1	0·46	0·18
Fat cover over side (score)	3·7	0·82	0·04	0·18	0·11
Fat thickness (mm)	5·3	4·70	0·05	0·04	0·11
Offal					
Caul fat as % carcass wt	2·36	0·276	0·012	0·17	0·11
Feet as % carcass wt	3·57	0·042	0·005	0·40	0·17
Hide as % live weight	6·66	0·191	0·027	0·50	0·19

Based on 46 d.f. between and 471 within progeny groups except:

a ,, ,, 50 ,, ,, ,, 557 ,, ,, ,,
b ,, ,, 45 ,, ,, ,, 536 ,, ,, ,,
c ,, ,, 41 ,, ,, ,, 493 ,, ,, ,,
d ,, ,, 32 ,, ,, ,, 390 ,, ,, ,,
e ,, ,, 53 ,, ,, ,, 609 ,, ,, ,,

for the butterfat variance, but since averages of the daughters' butterfat percentages were used the weights are the actual numbers of daughters. The second term in the formula is a correction due to the fact that the milk yield contemporary comparisons and butterfat averages are only estimates of the true breeding values. Values of $72{\cdot}375 \times 10^4$ kg^2 (33 094 gal^2) and 0·076, taken from the results of Barker and Robertson (1966), were used as estimates of the within-bull milk and butterfat variances. Correlations were calculated from the variances and covariances in the usual manner.

Since formulae for the standard errors of the milk-beef correlations were not available, empirical standard errors were calculated using the jack-knife technique of Tukey (1958).

RESULTS

The characters studied, together with their means, variances and heritability estimates, are listed in Table 1. A further set of heritabilities were calculated within three subsets of the data, namely progeny groups slaughtered at constant age, those slaughtered at constant weight and those so managed that they were not fit for slaughter till an age of over 25 months. The two age-groups were not, of course, mutually exclusive. The results are listed in Table 2. Characters were selected for this analysis according to the

TABLE 2

Heritability estimates (± SE) within subsets slaughtered according to different criteria or kept under similar management

Carcass character	Slaughtered at constant weight		Slaughtered at constant age		Kept to an age of 2–2½ yr	
	h^2	SE	h^2	SE	h^2	SE
Weight for age	0·33	0·38	0·25	0·22	0·32	0·25
Dressing %	0·59	0·36	0·46	0·34	0·54	0·42
Grade	0·25	0·26	0·25	0·19	0·21	0·14
Fleshing index	0·26	0·21	0·12	0·21	0·21	0·18
Ranginess	0·37	0·18	0·65	0·30	0·61	0·35
Blockiness of leg	0·28	0·40	0·49	0·25	0·53	0·33
Fat thickness	−0·02	0·11	−0·02	0·13	0·14	0·19

TABLE 3

Variation in groups slaughtered according to different criteria or kept under similar management. (Standard deviations within centre-years)

	Slaughtered at constant weight		Slaughtered at constant age		Kept to an age of 2–2½ yr	
	Mean	SD	Mean	SD	Mean	SD
Age at slaughter (days)	622	27·5	693	7·7	838	19·4
Live weight (kg)	448	17·9	477	32·3	511	34·9
Carcass weight (kg)	252	10·7	274	21·0	299	23·5
Carcass weight/age (kg/day)	0·411	0·027	0·405	0·031	0·405	0·029
Dressing %	56·2	1·5	57·4	1·9	58·3	1·8
Carcass grade (score)	4·3	1·6	5·6	1·55	5·0	1·73
Fleshing index (kg/m)	200	7·7	215	13·2	228	13·7
Ranginess (depth as % leg)	48·8	1·27	48·2	1·37	49·2	1·32
Blockiness	34·0	1·64	34·7	1·60	33·8	1·69
Fat thickness (mm)	4·5	2·0	6·3	2·4	5·0	2·1

TABLE 4

Correlations among carcass characters (phenotypic below diagonal, genetic above)

	Weight for age	Dressing %	Grade	Fleshing index	Ranginess	Blockiness of leg	Plumpness of leg	Eye-muscle area	% muscle in rib	Muscle/bone ratio	% fat in rib	Fat cover over side	Caul fat as % carcass	% bone in rib	Feet as % carcass	Hide as % live weight
Carcass weight for age		0·13	0·72	0·87	−0·78	0·96	0·49	0·60	−0·33	0·00	0·52	0·81	0·66	−0·76	−1·09	−0·18
Dressing %	0·32		0·57	0·48	−0·50	0·37	0·91	0·77	−0·28	−0·20	0·33	0·16	−0·12	−0·18	−0·55	0·00
Grade	0·25	0·31		0·88	−0·15	0·28	0·91	0·05	−0·45	−0·37	0·51	0·80	0·64	−0·16	−0·23	−0·16
Fleshing index	0·84	0·49	0·41		−0·99	0·70	0·90	0·85	−0·16	0·00	0·28	0·50	0·59	−0·36	−0·99	−0·21
Ranginess	−0·12	−0·09	−0·02	0·00		−0·51	−0·68	−0·74	−0·24	−0·56	0·03	−0·35	−0·58	0·73	0·36	0·05
Blockiness of leg	0·20	0·12	0·24	0·25	−0·06		0·72	0·31	−0·09	−0·10	0·06	0·66	0·27	−0·09	−0·14	0·27
Plumpness of leg	0·27	0·22	0·48	0·35	−0·09	0·28		0·48	−0·02	−0·13	0·06	0·51	0·14	0·20	−0·14	0·16
Eye-muscle area	0·28	0·26	0·15	0·36	−0·11	0·08	0·17		0·80	1·19	−0·71	−0·55	0·55	−0·61	−0·48	−0·20
% muscle in rib	−0·07	−0·10	−0·20	−0·09	−0·12	−0·11	−0·09	0·19		0·89	−0·93	−0·40	0·21	0·29	−0·03	−0·43
Muscle/bone ratio	0·13	0·10	0·07	0·18	0·04	0·04	0·09	0·21	0·61		−0·65	0·42	0·87	−0·75	−0·26	−0·48
% fat in rib	0·18	0·18	0·30	0·24	0·12	0·16	0·17	−0·11	−0·88	−0·24		0·47	0·09	−0·41	−0·16	0·35
Fat cover over side	0·21	0·15	0·56	0·23	0·01	0·18	0·34	−0·04	−0·32	−0·03	0·41		0·62	−0·24	−0·34	0·19
Caul fat (% carcass)	0·03	−0·04	0·03	0·07	0·06	0·10	0·04	−0·06	−0·21	0·02	0·27	0·10		−0·19	−0·07	0·05
% bone in rib	−0·23	−0·20	−0·26	−0·31	−0·03	−0·14	−0·19	−0·12	0·00	−0·19	−0·62	0·19	−1·09		0·67	0·06
Feet (% carcass)	−0·37	−0·39	−0·27	−0·50	0·01	−0·22	−0·23	−0·23	0·10	−0·20	−0·24	−0·19	0·21	0·33		0·52
Hide (% live wt)	−0·13	0·04	−0·26	−0·13	−0·05	−0·06	−0·06	−0·04	−0·12	0·12	0·06	−0·02	−0·10	0·05	0·25	

r_g based on 46 d.f. between and 472 d.f. within progeny groups.
$SE r_p = \pm 0{\cdot}046$. $SE r_g$ mostly lie between 0·25 and 0·55.

number of records available. Table 3 shows the extent to which we were successful in having steers slaughtered at constant weight or age and the effect this had on variation in other criteria (including fatness).

The phenotypic and genetic correlations among the carcass characters are set out in Table 4. Genetic correlations involving fat thickness are not meaningful because of its low heritability and have been omitted. For weight for age, grade and fleshing index correlations were also calculated for the slightly larger sample on which the heritabilities of these characters listed in Table 1 are based. For the characters in Table 2 further correlations were calculated within the age-constant, weight-constant and high-age subsets. In all three cases the phenotypic correlations were almost identical with those in Table 4 and the genetic correlations were not sufficiently different to be worth quoting.

The progeny test for milk (contemporary comparison) and the mean butterfat test of the daughters of the 66 bulls were taken from the latest figures in the Milk Marketing Board's published lists of *Friesian Bulls with Contemporary Comparisons*. Contemporary comparisons ranged from −360 to +522 kg and mean butterfat percentage from 3·46 to 4·02. The average effective number of daughters (the weighting) was 320.

The genetic correlations between beef and milk characters are shown in Table 5. These are based on the same animals as in Table 1 for the live

TABLE 5

Genetic correlations between milk and beef characters in progeny of AI bulls

Beef characters of steers	Genetic correlations with heifers' milk yield (contemporary comparison)	butterfat %
Live†		
Gain to 6 mo	0·31	0·18
Gain 6–12 mo	−0·47	0·19
Withers height	−0·03	0·11
Chest depth	0·16	−0·63
Carcass		
Weight for age	−0·68	−0·09
Dressing %	0·01	−0·58
Grade	0·15	−0·21
Fleshing index	−0·59	−0·51
Ranginess	0·97	−0·08
Blockiness of leg	−0·47	−0·42
Plumpness of leg	−0·24	−0·32
Eye-muscle area	−0·21	−0·59
% muscle in rib	−0·23	−0·01
Muscle/bone ratio	−0·40	−0·06
% fat in rib	−0·08	−0·06
Fat cover over side	−0·55	−0·10
Caul fat as % carcass	−0·68	−0·44
% bone in rib	0·61	−0·18
Feet as % carcass	0·43	0·45
Hide as % live weight	0·01	0·03

† SEr_g lie between 0·27 and 0·65.

animal characters and as Table 3 for the carcass characters. When the correlations were calculated within the subsets of similar management the results shown in Table 6 were obtained.

TABLE 6

Genetic correlations between milk yield of heifer progeny of AI bulls and some carcass characters of steer progeny slaughtered at constant age, at constant weight and at an age of over 2 years ($r_g \pm SE_{r_g}$)

Carcass character	Constant weight		Constant age		Slaughtered at 2–2½ yr	
	r	SE	*r*	SE	*r*	SE
Weight for age	−0·72	0·19	−0·92	1·06	0·03	0·49
Dressing %	−0·44	0·50	0·43	0·23	0·17	0·44
Grade	0·22	0·82	−0·33	0·54	0·12	0·48
Fleshing index	−0·56	0·36	−0·88	0·45	0·01	0·59
Ranginess	1·03	0·31	0·87	0·20	0·57	0·39
Blockiness of leg	−0·64	0·34	−0·51	0·44	0·20	0·38

DISCUSSION

Heritabilities. For most of the characters listed in Table 1 heritabilities had been calculated previously from some of the first progeny groups slaughtered and were listed in a preliminary report (Vial and Mason, 1962). That analysis also included some carcass characters, e.g. cannon bone dimensions and marbling of eye muscle, for which there were insufficient additional observations to justify recalculation of heritabilities. For most traits the preliminary heritability estimates were not significantly different from the present ones; exceptions are carcass depth/length (ranginess) for which the estimate has risen from −0·13 to +0·50, and percentage of muscle in rib joint (0·04 to 0·58).

It might be argued that these estimates are invalid because of the different management conditions and different slaughter criteria for the various centre-years. It was in an attempt to investigate the effect of these differences that the analysis was made within subsets of the data with the same slaughter criteria (age or weight) or with management conditions adjusted to give a lower growth rate and therefore a high age at slaughter. The results (Table 2) show that, in relation to their standard errors, the heritabilities vary little between the three subsets of the data. In fact they vary less than between the preliminary and final analyses. There is thus no evidence of interaction between method of management or criterion for slaughter and heritability.

The heritabilities of rate of gain to 6 months and from 6 to 12 months are similar to those listed in the review by Preston and Willis (1970) for pre-weaning growth rate of beef cattle for which the median value of 35 estimates is 0·27. Figures for Friesians in the literature are 0·12 for heritability of gain to 9 months (Hodges, O'Connor and Higgin, 1961) and 0·33 for weight for age at 12 months (Soller, Shilo and Bar-Anan, 1966). The reason for our estimate of zero heritability of gain from 12 to 18 months is not clear; it does not accord with the estimate of 0·34 for live weight for age at slaughter. This is similar to other figures for Friesians, e.g. 0·45 (Langlet, Gravert and Rosenhahn, 1967), 0·26 (Cunningham and Broderick, 1969). For carcass weight

for age our estimate of heritability (0·18) is rather lower than the four quoted by Preston and Willis (1970) which range from 0·28 to 0·49; three of them refer to dairy breeds.

Our heritability estimates are highest for measurable aspects of carcass conformation (ranginess, blockiness, weight of feet), for quantitative measures of muscle and fat (eye-muscle area, proportions of muscle and fat in the rib joint), and for weight of hide. These then are the characters which can be changed most easily by selection—which accords with the known breed differences, even within the black-and-white (Friesian) group of breeds, in length of leg, earliness of maturity and skin thickness. Preston and Willis (1970) also list high heritabilities for eye-muscle area (in many cases, like ours, not corrected for carcass weight) and our high figures for carcass grade and dressing percentage accord with the figures they quote. On the other hand they quote a low heritability for percentage of muscle in the rib joint and a high one for percentage of bone; Cunningham and Broderick (1969) also found, in both Friesians and Shorthorns, a low heritability estimate for muscle percentage and a high one for bone. Our low heritability for bone in rib is associated with its low genetic variability and the inaccuracy of the cutting out of the rib which will affect flesh/bone ratio much more than muscle/fat ratio (where flesh = muscle+fat).

The low heritability of subcutaneous fat thickness compared with the other measures of fatness may be due to the extreme difficulty of accurately measuring thin layers of fat in unfrozen carcasses. It was higher (0·28) when the estimate was based on the average of five measurements (Vial and Mason, 1962) instead of the present three. Preston and Willis (1970) list estimates ranging from 0·24 to 0·74.

Correlations among beef traits. The highest phenotypic correlation was 0·91 between live weight for age and carcass weight for age which compares with the 0·90 of Cunningham and Broderick (1969). Because of this close relationship the live-weight correlations have not been quoted—they repeat the correlations with carcass weight. Most of the phenotypic correlations in Table 4 are significantly different from zero (because they are based on large numbers) but not high enough to be of any predictive value. The only high ones are 0·84 between fleshing index and carcass weight for age and −0·88 between the percentages of fat and muscle in the rib joint. The first is only to be expected because carcass weight is the chief component of both variables. The second arises from the fact that bone percentage varies little so that high fat must mean low muscle percentage. The muscle/bone ratio is much less affected by fatness than is the muscle percentage itself, which is the argument for using this ratio as an index of carcass muscling. However, its heritability appears to be lower than that of percentage muscle. As predictors of eye-muscle area, of muscle percentage or of muscle/bone ratio, which are the important measures of muscling in the carcass, the various indexes of carcass fleshing, grade or conformation of the uncut carcass, are no better than carcass weight for age. Fat cover over the side and grade are fairly highly correlated with percentage fat in the rib but not high enough to be good predictors. The correlations involving grade show clearly that leg plumpness and external fat cover were important factors influencing the Ministry of Agriculture inspectors in assessing carcasses.

The genetic correlations appear high but their standard errors are also high so that their actual values must not be taken too seriously. Those which

are highest suggest that selecting for growth rate will increase the desirability of carcass conformation and will do this by reducing the length of leg and increasing the fatness of the carcass rather than by increasing size and muscling. This is rather surprising since it is usually assumed that fatness (earliness of maturity) is associated with small size and slow growth rate. Certainly this applies when comparisons are made between breeds (*see* Mason, 1971) but within the breeds, the few results are conflicting. On *ad libitum* feeding in U.S.A. genetic correlations calculated between growth rate and carcass fatness have been positive (Shelby, Harvey, Clark, Quesenberry and Woodward, 1963; Cundiff, Chambers, Stephens and Willham, 1964). On the other hand, on pasture in New Zealand, Brumby, Walker and Gallagher (1963) calculated a negative genetic correlation between live-weight gain and fat cover of the carcass of Aberdeen-Angus cattle. As to the results of selection for growth rate Gallagher (1964) showed that in the New Zealand experiment a line selected for fast growth produced leaner carcasses than one selected for slow growth in the first year when animals were slaughtered at 290 kg live weight and 18 to 21 months of age; in later years animals were slaughtered at a constant (and annually decreasing) age of 754 to 603 days and during these years there was no difference between the two lines in fat cover or marbling. In America (Front Royal, Virginia, Shorthorn and Angus lines selected for rapid growth between weaning and yearling ages appear so far to have either changed little in fatness or have become less fat (R. C. Carter and K. P. Bovard, personal communications, 1971). These are preliminary reports on small numbers. In a Canadian selection experiment for yearling weight in Shorthorn bulls the weight of the selected line was greater than the control by 29 kg at 419 days but carcass composition was not affected except that kidney fat was slightly greater (Martin, Fredeen and Newman, 1970).

Milk-beef correlations. The genetic correlations in Tables 5 and 6 are much greater, arithmetically, than any previously published but they also have high standard errors. The review of literature in Mason (1964) and the results of later authors (e.g. Wilk, Young and Cole, 1963; Dinklage, 1965; McDaniel and Legates, 1965; Martin and Starkenburg, 1965; Samson-Himmelstjerna, 1965; Lefebvre and Ricordeau, 1966; Soller, Shilo and Bar-Anan, 1966; Syrstad, 1966; Langlet, Gravert and Rosenhahn, 1967; Rønningen, 1967; Jesswein, 1968; Brum and Ludwick, 1969) suggested that the genetic correlations between milk yield and growth rate, body weights and measurements, whether measured on cows or steers, were positive but small and, in most cases, not significantly different from zero. However, some more recent results (Suess, Tyler and Brungardt, 1968; Temisan, Constantinescu and Oltenacu, 1970; Bar-Anan, 1971) show significant negative correlations between measures of growth of a sire's progeny and his milk-yield progeny test. The difference between the result of Bar-Anan (1971) and the zero correlation obtained by Soller *et al.*, (1966), both on Israeli Friesians, is probably explained by the decreasing age at first calving in Israel. Bar-Anan (personal communication, 1971) showed that the growth rate of the bull's sons was positively correlated with a milk progeny test based on daughters first calving at over 2 yr of age but negatively with one based on daughters calving at younger ages. He interprets this to mean that the rapidly growing bulls produce daughters which are late maturing in respect of milk yield just as the bulls themselves (and their progeny) are in respect to adult size.

Our figures also suggest that the genetic correlations between growth and

milk yield are large and negative but in connection with Bar-Anan's observation on the effect of precocity, it may be significant that these negative signs disappear when the correlations are based on older progeny. (Table 6). If only the high correlations in Table 5 are heeded it would appear that selection solely for milk yield within the Friesian breed would produce a slow growing animal (low weight for age), with a poor conformation (carcass deep and light in relation to its length), with little aptitude to fatten and with much bone (measured here in rib and feet). This is curiously similar to the derogatory description of the dairy beast given by the breeder of specialized beef cattle.

However, if the Friesian is compared with the British beef breeds (e.g. Callow, 1961; Cole, Ramsey, Hobbs and Temple, 1964; Royal Smithfield Club, Agricultural Research Council and Norfolk Agricultural Research Station, 1966) while it is indeed slightly inferior in muscle/bone ratio and has less fat (although this is no longer a disadvantage), it is actually superior in growth rate, in proportion of muscular tissue and in proportion of high price cuts. On the other hand, compared with the large continental breeds (*see* Mason, 1971) some of the differences are reversed. The Friesian now shows up as smaller and fatter but still with a lower muscle/bone ratio. If any generalization is to be made about the genetic difference in beef traits between beef and dairy breeds it may be that it will lie in the high proportion of bone in dairy breeds. However, more carefully controlled experiments on large numbers of animals will be needed to confirm, within a breed, a high positive correlation between percentage of bone and milk yield.

ACKNOWLEDGEMENTS

This research was organized and sponsored by the Agricultural Research Council with the cooperation of many individuals and institutions.

We wish to thank the headquarters and field staff of the Milk Marketing Board (England and Wales), Mr H. J. Ballinger (A.R.C. Cambridge), and the staff of the Cattle Breeding Centre, Shinfield, Reading, for help in the location and collection of calves. We are most grateful to the Fatstock Marketing Corporation for supplying facilities for carcass examination at slaughterhouses. We are deeply indebted to the following cooperators who reared steers: Frederick Hyam Ltd, Cambridge; School of Agriculture, University of Newcastle upon Tyne; Yorkshire (WR) Institute of Agriculture, Askham Bryan; Norfolk Agricultural Station, Sprowston; British Oil and Cake Mills Ltd, Selby, Yorks; R. Silcock and Sons Ltd, Tushingham, Shropshire; Spillers Ltd, Middle Aston, Oxford; N.A.A.S. Gleadthorpe Experimental Husbandry Farm, Mansfield, Notts.

Above all our thanks are due to Professor M. McG. Cooper and his staff at the Newcastle School of Agriculture. One of us (V. E. V.) was working there at the time of this experiment; he collected and recorded a large number of the experimental animals. When he left Mr K. W. Leach and Mrs Anne Black assisted in the examination of carcasses and the dissection of rib joints, respectively.

REFERENCES

AGRICULTURAL RESEARCH COUNCIL. 1965. Recommended procedures for use in the measurement of beef cattle and carcasses. *Agricultural Research Council, London.* 18 pp.

BAR-ANAN, R. 1971. [Some problems in breeding dual-purpose cattle.] *Züchtungskunde* **43**: 74–76.

BARKER, J. S. F. and ROBERTSON, ALAN. 1966. Genetic and phenotypic parameters for the first three lactations in Friesian cows. *Anim. Prod.* **8**: 221–240.

BRUM, E. W. and LUDWICK, T. M. 1969. Heritabilities of certain immature and mature body measurements and their correlations with first-lactation production of Holstein cows. *J. Dairy Sci.* **52**: 352–359.

BRUMBY, P. J., WALKER, D. K. and GALLAGHER, R. M. 1963. Factors associated with growth in beef cattle. *N.Z. Jl agric. Res.* **6**: 526–537.

CALLOW, E. H. 1961. Comparative studies of meat. VII. A comparison between Hereford, Dairy Shorthorn and Friesian steers on four levels of nutrition. *J. agric. Sci., Camb.* **56**: 265–282.

COLE, J. W., RAMSEY, C. B., HOBBS, C. S. and TEMPLE, R. S. 1964. Effects of type and breed of British, Zebu and dairy cattle on production, carcass composition, and palatability. *J. Dairy Sci.* **47**: 1138–1144.

CUNDIFF, L. V., CHAMBERS, D., STEPHENS, D. F. and WILLHAM, R. L. 1964. Genetic analysis of some growth and carcass traits in beef cattle. *J. Anim. Sci.* **23**: 1133–1138.

CUNNINGHAM, E. P. and BRODERICK, T. 1969. Genetic and environmental parameters of growth and carcass traits in dual-purpose cattle. *Ir. J. agric. Res.* **8**: 397–416.

DINKLAGE, H. 1965. [Relationship between milk, fat content and meat producing capacity of German Simmental cattle in Bavaria.] *Bayer. landw. Jb.* **42**: 702–720.

GALLAGHER, R. M. 1964. The influence of growth rate selection on some carcass characteristics of beef cattle. In '*Carcase Composition and Appraisal of Meat Animals*': *Tech. Conf., East Melbourne* (ed. D. E. Tribe), pp. 9–1 to 9–9.

HODGES, J., O'CONNOR, L. K. and HIGGIN, R. 1961. The growth and size of Friesian cattle on commercial farms in England. *8th int. Congr. Anim. Prod., Hamburg*, 1961 **2**: 30–31.

JESSWEIN, H. 1968. [The association between milk production, fattening performance and carcass composition in German Black Pied cattle.] *Züchtungskunde* **40**: 197–204.

LANGLET, J. F., GRAVERT, H. O. and ROSENHAHN, E. 1967. [Investigations on the heritability of beef production in Black Pied cattle.] *Z. Tierzücht. ZüchtBiol.* **83**: 358–370.

LEDGER, H. P. and HUTCHISON, H. G. 1962. The value of the tenth rib as a sample joint for the estimation of lean, fat and bone in carcasses of East African Zebu cattle. *J. agric. Sci., Camb.* **58**: 81–88.

LEFEBVRE, J. and RICORDEAU, G. 1966. Genetic correlations between milk production and conformation in the Montbéliard and Normandy cattle breeds. *9th int. Cong. Anim. Prod., Edinb., Scient. Progrm. Abstr.*, Eng. ed: p. 16 (Abstr.).

MCDANIEL, B. T. and LEGATES, J. E. 1965. Association between body weight predicted from heart girth and production. *J. Dairy Sci.* **48**: 947–956.

MARTIN, A. H., FREDEEN, H. T. and NEWMAN, J. A. 1970. Live performance and carcass characteristics of beef bulls in relation to selection for yearling weight. *Can. J. Anim. Sci.* **50**: 491–497.

MARTIN, T. G. and STARKENBURG, R. T. 1965. Genetic correlations between beef and dairy traits in dual purpose cattle. *Wld Rev. Anim. Prod.* **1** (Special issue): 45–52.

MASON, I. L. 1951. Performance recording in beef cattle. *Anim. Breed. Abstr.* **19**: 1–24.

MASON, I. L. 1964. Genetic relations between milk and beef characters in dual-purpose cattle breeds. *Anim. Prod.* **6**: 31-45.

MASON, I. L. 1971. Comparative beef performance of the large cattle breeds of western Europe. *Anim. Breed. Abstr.* **39**: 1–29.

PRESTON, T. R. and WILLIS, M. B. 1970. *Intensive Beef Production.* Pergamon Press, Oxford.

ROBERTSON, ALAN, STEWART, A. and ASHTON, E. D. 1956. The progeny assessment of dairy sires for milk: the use of contemporary comparisons. *Proc. Br. Soc. Anim. Prod.*, 1956, pp. 43–50.

RØNNINGEN, K. 1967. Phenotypic and genetic parameters for characters relating to milk production in cattle. *Acta Agric. scand.* **17**: 83–100.

ROYAL SMITHFIELD CLUB, AGRICULTURAL RESEARCH COUNCIL AND NORFOLK AGRICULTURAL RESEARCH STATION. 1966. *Major beef research project. A comparison of the growth of different types of cattle for beef production.* Underhill (Plymouth) Ltd. 53 pp.

SAMSON-HIMMELSTJERNA, D. VON. 1965. [Correlations between milk yield, fat percentage and weight gain in Oldenburg Black Pied cattle.] *Züchtungskunde* **37**: 323–330.

SEARLE, S. R. 1961. Variance components in the unbalanced 2-way nested classification. *Ann. math. Stat.* **32**: 1161–1168.

SHELBY, C. E., HARVEY, W. R., CLARK, R. T., QUESENBERRY, J. R. and WOODWARD, R. R. 1963. Estimates of phenotypic and genetic parameters in ten years of Miles City ROP steer data. *J. Anim. Sci.* **22**: 346–353.

SOLLER, M., SHILO, A. and BAR-ANAN, R. 1966. A note on heritability of live-weight-for-age in Israel-Friesian dairy bull calves and its genetic correlation with milk yield. *Anim. Prod.* **8**: 157–158.

SUESS, G. G., TYLER, W. J. and BRUNGARDT, V. H. 1968. Relationship between carcass characteristics of Holstein steers and genetic level of milk production. *J. Anim. Sci.* **27**: 972–975.

SYRSTAD, O. 1966. Studies on dairy herd records. IV. Estimates of phenotypic and genetic parameters. *Acta. Agric. scand.* **16**: 79–96.

TALLIS, G. M. 1959. Sampling errors of genetic correlations coefficients calculated from analyses of variance and covariance. *Aust. J. Stat.* **1**: 35–43.

TEMISAN, V., CONSTANTINESCU, C. and OLTENACU, A. 1970. [Phenotypic and genetic correlations between milk and meat production in Romanian Simmental cattle]. *Lucr. stiint. Inst. Cerc. zooteh.*, **27**: 73–84.

TUKEY, J. W., 1958. Bias and confidence in not-quite large samples. *Ann. math. Stat.* **29**: 614 (Abstr.)

VIAL, V. E. and MASON, I. L. 1962. Heritability estimates for growth and carcase characters in Friesian steers. *8th int. Congr. Anim. Prod., Hamburg*, 1961 **3**: 20–23.

WILK, J. C., YOUNG, C. W. and COLE, C. L. 1963. Genetic and phenotypic relationships between certain body measurements and first lactation milk production in dairy cattle. *J. Dairy Sci.* **46**: 1273–1277.

(*Received* 15 *July* 1971)

38

Heritability of milk yield and composition at different levels and variability of production

by

W.G. Hill, M.R. Edwards and M-K A. Ahmed and R. Thompson

Anim. Prod. 1983, **36**: 59-68
© 1983 British Society of Animal Production
0003-3561/83/22610059$02-00

HERITABILITY OF MILK YIELD AND COMPOSITION AT DIFFERENT LEVELS AND VARIABILITY OF PRODUCTION

W. G. HILL, M. R. EDWARDS, M.-K. A. AHMED†
Institute of Animal Genetics, University of Edinburgh, West Mains Road, Edinburgh EH9 3JN

AND

R. THOMPSON
ARC Unit of Statistics, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ

ABSTRACT

Analyses of variance were conducted on first lactation milk, fat and protein production records in England and Wales of daughters of British Friesian sires. Herds were split on milk yield into high and low levels of mean production and, in subsequent analyses, into high and low levels of within herd variance and coefficient of variation using all first lactation records. Data were then extracted on daughters of 798 young sires undergoing progeny test and on 118 widely used proven sires to generate connections. Least squares analyses were conducted within levels and genetic correlations estimated from the covariance of sire effects.

With data split on mean yield, the heritability of milk yield was 0·24 at the low level and 0·30 at the high level, that of log transformed yield being 0·25 and 0·35 respectively. With data split on variance the corresponding figures were 0·24, 0·30, 0·27 and 0·36 respectively, and when split on coefficient of variation, 0·22, 0·26, 0·26 and 0·32. There were similar increases for fat and protein yield, proportionately smaller increases for fat and protein content.

Genetic correlations were close to 1·0 between high and low levels for all traits on all criteria of data splitting.

As a consequence progeny testing of bulls is rather more accurate at high mean or variance of production levels and data can be combined optimally without scaling. Cows of the highest predicted value using an index will be found in high variance herds.

INTRODUCTION

EVALUATION of dairy sires from field records involves combining data from herds with different management systems. This raises two main questions: are tests in different systems equally accurate and do sires rank the same in each? In the analysis reported in this paper herds were divided into two classes according to mean level of production. Although production level does not specify any single management scheme, herds can be classified from the production records alone; and in similar previous studies higher heritabilities have usually been obtained at higher production levels (Maijala and Hanna, 1974), although not in British data (Robertson, O'Connor and Edwards, 1960). In high-producing herds, variance of yield is also higher, so herds were also divided by variance and coefficient of variation of yield to determine, if possible, whether the primary cause of increased heritability was mean or variance of production. The results have implication both for sire evaluation and for indexing of individual cows among which, for bull dam selection, comparisons have to be made between different herds.

MATERIAL AND METHODS

Records of first lactation production on progeny of Friesian-Holstein sires (bulls) were kindly supplied by the Milk Marketing Board of England and Wales (MMB). The data comprised 305-day records (if the lactation was complete, otherwise at

† Present address: Department of Animal Production, Faculty of Agriculture, University of Khartoum, Shambat, Sudan.

least 200-day records) of milk, fat and protein yields and of fat and protein contents. Logarithms of yields were also computed and analysed.

Herds were extracted which had daughters of bulls entering the MMB's unproven stud in the years 1968/9 to 1975/6. These young (unproven) bulls were progeny tested in one of five regions, and were classified by year of entering test and region into 36 groups. Bulls whose progeny tests were not completed within the years of data available were discarded. The groups were put into 24 sets, with one or two successive complete groups of sires from one region comprising a set, the division into sets being necessary because the number of effects which could be included in each least squares analysis was limited. For each set of sires, all records were extracted for each year on herds having any daughters of a young sire in the set calving in that year.

Split by mean. These herds were then split into two groups on the basis of the mean milk yield of all females in their first lactation (heifers) (by the young and all other sires), corrected for age and month of calving using standard MMB correction factors for sire data (G. J. T. Swanson, personal communication) over the 3- to 4-year period during which the set of young sires had daughters calving. The cut-off point was chosen so that there were nearly equal numbers of young sires' progeny in each subdivision. The records of daughters of the young sires were then extracted, together with those on 15 to 20 of the most widely used old sires in these herds, which were taken to help generate connexions for estimating sire effects because the number of daughters of young sires in any herd-year-season was small. The extracted data comprised about one-third (0·31) of that on which the split on herd mean was made. Some daughters of old sires appeared in more than one set of data if they were contemporaneous to daughters of different groups of young sires. The data structure is outlined in Table 1; overall there were included nearly 27 000 daughters of 798 young sires and nearly 134 000 records on 118 old sires, with rather more than one-third (0·36) of the latter being duplicates.

Least squares analyses using the program of Harvey (1977) were conducted at each level of the 24 sets, making 48 analyses in all. The model used to describe an individual record was

$$Y_{ijklm} = \mu + H_i + G_j + S_{jk} + T_l + b_1 D + b_2 D^2 + b_3 D^3 + b_1' A + b_2' A^2 + e_{ijklm} \qquad (1)$$

where: μ = overall mean; H_i = effect of ith herd-year-season of calving; G_j = effect of jth group of sires ($j = 1$ denotes old sires, $j = 2, 3$ denotes young

TABLE 1
Structure of data

Groups of young sires	36
No. of analyses	24
No. of young sires	798 (797†)
No. of old sires	118
Duplication of daughters of old sires	0.363
Mean no. old sires per analysis	18.8

Criterion for splitting	Unsplit	Mean		Variance		CV	
Level		Low	High	Low	High	Low	High
No. daughters young sires	26 710	13 646	13 064	13 559	13 151	13 308	13 390
No. daughters old sires‡	133 814	52 483	81 331	57 183	76 631	65 219	68 595
d.f. within sires§	126 738	50 396	74 766	53 798	71 364	60 368	64 785
d.f. among young sires	762	762	762	762	762	761†	761†
Coefficient young sire component in e.m.s. (k)	22·94	11·37	11·35	11·27	11·45	11·20	11·55

† One sire omitted from data split by CV because all daughters at one level.
‡ Includes duplicate records.
§ d.f. within sires for unsplit data exceeds the sum in split data sets by d.f. fitted for fixed effects in each analysis.

sires); S_{jk} = effect of kth sire in jth group; T_l = effect of lth month of calving; b_{1_1}, b_2, b_3 are regressions on length of lactation, D; b_1', b_2' are regressions on age at calving, A, with D and A expressed as deviations from herd-year-season mean; and e_{ijklm} = residual error. The residual variance, essentially the variance within sires, $V(e) = \sigma_W^2$, was computed from progeny of both young and old sires; the variance between sires, $V(S) = \sigma_B^2$, was computed solely within groups of young sires.

To obtain a pooled estimate of σ_W^2, the residual sums of squares were added from each set of data. The pooled estimate of σ_B^2 was obtained by weighting the estimate from each group of young sires by the inverse of its variance. Because there were so many d.f. within sires, the variance within sires was assumed to be estimated with trivially small sampling error relative to that between sires. The variance, $V(\hat{\sigma}_{B_j}^2)$, of the estimate of sire variance from the jth group was therefore taken as

$$V\hat{\sigma}_{B_j}^2) = 2(\sigma_W^2 + K_j\sigma_B^2)^2 / (d_j K_j^2) \qquad (2)$$

where K_j is the coefficient of σ_B^2 in the expected mean square for group j of young sires and d_j are the corresponding degrees of freedom. An iterative procedure was used to obtain the pooled estimate (Nelder and Wedderburn, 1972).

The genetic covariances between performance at high and low production levels were computed from the weighted cross products of the estimates of effects of young sires, $\hat{S}_{Hjk}$ and $\hat{S}_{Ljk}$ at high and low levels, respectively. The weights used were inversely proportional to

$$V(Cov(\hat{S}_{Hjk}, \hat{S}_{Ljk})) = Cov_B^2 + (\sigma_{HW}^2/n_{Hjk} + \sigma_{HB}^2)(\sigma_{LW}^2/n_{Ljk} + \sigma_{LB}^2) \qquad (3)$$

where n_{Hjk}, n_{Ljk} are the 'effective numbers' of progeny of the sire (i.e. inversely proportional to the variance of the estimate of the sire effect) computed in the corresponding least squares analysis and Cov_B is the covariance between high and low sire effects. The estimates of variance were taken from the individual analyses and the covariance was estimated by iteration, subject to the constraint that $\hat{Cov}_B^2 \leq \sigma_{HB}^2\sigma_{LB}^2$ in the weighting formula (3). The genetic correlation was estimated as $\hat{r}_G = \hat{Cov}_B/(\hat{\sigma}_{HB}\hat{\sigma}_{LB})$.

Standard errors of intra-class correlations and genetic correlations were obtained, by assuming the variances within sires were known without error, using standard methods (e.g. Guiard and Herrendörfer, 1977); in view of the approximations and complexities involved in estimating s.e.$(\hat{r}_G)$, the weighting structure was ignored in its derivation.

Split by variance and coefficient of variation (CV). The data were also split into two groups using as a criterion the pooled variance of milk yield within year-seasons for each herd over the same 3- to 4-year period and using all heifer records, as for the split by mean, and the same analysis was repeated. The data were split for a third time by the within year-season coefficient of variation using the variances and mean computed for the previous splits, and the analysis repeated, except that one sire family was eliminated (Table 1) because all heifers were in herds of the same category for CV. The distribution of records in different categories are shown in Table 2. For example, about one-quarter (0·273) of records were in the high mean, high variance and high CV category, and about two-thirds (0·662) of records were in the high mean-high variance or low mean-low variance category.

TABLE 2

Proportions of records in each category after splitting by mean, variance and CV

		Low variance	High variance	Total
Low mean	Low CV	0·172	0·001	0·172
	High CV	0·084	0·154	0·239
	Total	0·256	0·155	0·411
High mean	Low CV	0·182	0·134	0·316
	High CV	0·000	0·273	0·273
	Total	0·182	0·407	0·589
Total	Low CV	0·353	0·135	0·488
	High CV	0·085	0·427	0·512
	Total	0·438	0·562	1·000

Proportions in high-high and low-low categories

Mean-variance	0·662
Mean-CV	0·446
Variance-CV	0·781

TABLE 3

Results of analysis on whole data set, taken in one group (unsplit) and split by mean, variance and coefficient of variation

Criterion for splitting	Unsplit	Mean		Variance		CV	
		Low	High	Low	High	Low	High
Mean							
Milk yield (kg)	4666	4156	5023	4419	4860	4724	4611
Fat yield (kg)	179·3	158·6	193·8	170·5	186·3	182·7	176·1
Protein yield (kg)	150·5	133·5	162·4	141·8	157·3	151·9	149·1
Fat content†	3·847	3·820	3·866	3·856	3·840	3·869	3·827
Protein content†	3·226	3·213	3·236	3·208	3·241	3·216	3·236
Log_{10} milk yield	3·658	3·609	3·693	3·636	3·676	3·666	3·652
Log_{10} fat yield	2·242	2·190	2·278	2·221	2·259	2·252	2·233
Log_{10} protein yield	2·167	2·116	2·203	2·142	2·186	2·173	2·162
CV (× 100)							
Milk yield	14·2	14·2	14·2	12·5	15·2	12·6	15·6
Fat yield	13·9	14·1	13·8	12·6	14·6	12·6	15·1
Protein yield	13·0	13·1	12·9	11·6	13·8	11·6	14·3
Fat content	8·1	8·0	8·3	7·8	8·5	7·9	8·4
Protein content	5·2	5·3	5·2	5·1	5·4	5·1	5·3
Log_{10} milk yield	1·7	1·7	1·7	1·5	1·9	1·5	1·9
Log_{10} fat yield	2·8	2·9	2·7	2·5	2·9	2·5	3·0
Log_{10} protein yield	2·7	2·8	2·6	2·4	2·9	2·3	3·0
Variance within sire families							
Milk yield	413860	328500	470290	287610	505390	336330	483700
Fat yield	585·1	469·2	661·6	438·8	691·3	503·2	658·6
Protein yield	364·7	292·2	412·5	255·7	443·6	295·6	426·6
Fat content	0·08780	0·08330	0·09085	0·08018	0·09348	0·08360	0·09162
Protein content	0·02471	0·02487	0·02456	0·02342	0·02560	0·02342	0·02583
Log_{10} milk yield	0·003657	0·003706	0·003612	0·002894	0·004202	0·002841	0·004375
Log_{10} fat yield	0·003576	0·003690	0·003487	0·002970	0·004009	0·002875	0·004196
Log_{10} protein yield	0·003159	0·003241	0·003092	0·002507	0·003623	0·002424	0·003894

Milk yield	27610	21370	38390	18050	41140	19430	33980
Fat yield	37·3	29·4	50·8	25·2	51·1	26·4	44·7
Protein yield	20·3	15·8	27·9	12·8	30·4	13·9	25·5
Fat content	0·01055	0·00959	0·01169	0·00943	0·01256	0·01064	0·01108
Protein content	0·00388	0·00358	0·00405	0·00339	0·00463	0·00379	0·00408
Log_{10} milk yield	0·000282	0·000251	0·000346	0·000206	0·000417	0·000199	0·000378
Log_{10} fat yield	0·000264	0·000241	0·000317	0·000197	0·000367	0·000187	0·000352
Log_{10} protein yield	0·000200	0·000178	0·000240	0·000147	0·000299	0·000139	0·000274
Heritability (± s.e.) (× 100)							
Milk yield	25·0 (2·0)	24·4 (2·8)	30·2 (3·0)	23·6 (2·8)	30·1 (3·0)	21·9 (2·7)	26·3 (2·8)
Fat yield	24·4 (2·0)	23·6 (2·8)	28·5 (2·9)	21·8 (2·7)	27·5 (2·9)	19·9 (2·6)	25·4 (2·8)
Protein yield	21·1 (1·9)	20·5 (2·6)	25·4 (2·8)	19·1 (2·6)	25·7 (2·8)	18·0 (2·5)	22·6 (2·7)
Fat content	42·9 (2·7)	41·3 (3·4)	45·6 (3·6)	42·1 (3·4)	47·4 (3·6)	45·2 (3·6)	43·1 (3·4)
Protein content	54·3 (3·1)	50·3 (3·7)	56·7 (3·9)	50·6 (3·7)	61·3 (4·0)	55·8 (3·9)	54·6 (3·8)
Log_{10} milk yield	28·6 (2·2)	25·4 (2·8)	35·0 (3·2)	26·6 (2·9)	36·1 (3·2)	26·2 (2·9)	31·8 (3·0)
Log_{10} fat yield	27·5 (2·1)	24·6 (2·8)	33·3 (3·1)	24·9 (2·8)	33·5 (3·1)	24·4 (2·8)	30·9 (3·0)
Log_{10} protein yield	23·8 (2·0)	20·8 (2·6)	28·8 (3·0)	22·1 (2·7)	30·5 (3·0)	21·7 (2·7)	26·9 (2·9)
h^2 *(high)* − h^2 *(low) (± s.e.) (× 100)*							
Milk yield	—	5·8 (4·0)		6·5 (4·0)		4·4 (3·8)	
Fat yield	—	4·9 (3·9)		5·8 (3·8)		5·5 (3·7)	
Protein yield	—	4·9 (3·8)		6·5 (3·7)		4·6 (3·6)	
Fat content	—	4·3 (4·8)		5·3 (4·9)		−2·0 (4·8)	
Protein content	—	6·4 (5·3)		10·7 (5·4)		−1·2 (5·3)	
Log_{10} milk yield	—	9·6 (4·2)		9·5 (4·2)		5·7 (4·1)	
Log_{10} fat yield	—	8·7 (4·1)		8·6 (4·1)		6·5 (4·1)	
Log_{10} protein yield	—	8·0 (3·9)		8·3 (4·0)		5·1 (3·9)	

† Fat content = (fat yield (kg)/milk yield (kg)) × 100
Protein content = (protein yield (kg)/milk yield (kg)) × 100

Unsplit data. The data sets were also analysed as a whole, i.e. unsplit, and the same analysis, except computation of genetic correlations, performed.

RESULTS

The main results are given in Table 3. On the unsplit data heritabilities were around 0·25 for milk and fat yield, rather lower for protein yield. Transformation of yields to logarithms increased heritabilities by about 0·03. The heritability of protein content (0·54) was, however, substantially higher than for fat content (0·43).

When herds were split by *mean* milk yield, the coefficient of variation of milk yield was similar at each level; the variance within sires increased by a factor of 1·43 from low to high level herds and that between sires by 1·80, leading to an increase in heritability from 0·24 at low level to 0·30 at high level. Changes in variance of fat and protein yield were similar. For fat and protein content there was also an increase, albeit a proportionately smaller one, from the low to the high level. For log transformed yields the variance within families was slightly lower at the high level but that between families substantially larger (by factors of 0·97 and 1·38 for log milk yield) leading to an increase from 0·25 to 0·35 in heritability of log milk yield.

When data were split by *variance* within herds there was an increase from low to high in mean and in variance of yield both within and between sires for yields, log transformed yields and content. These changes were such that the heritabilities in low variance herds were almost the same as in low mean herds, and increased in high variance herds to almost the same values as in high mean herds.

Finally, with data split by *coefficient of variation* within herds, the mean level of production was lower in the high CV herds. Yet the increase in variance within and between sires from low to high CV herds was almost as great as that from low to high variance herds, such that for yield traits, but not composition traits, a higher heritability was found in higher CV herds.

Genetic correlations between production in high and low level herds split by each criterion are shown in Table 4. There is little evidence that the true value of any of the correlations differs appreciably from 1·0.

TABLE 4

Genetic correlation (r_G) between performance in herds split by mean, variance and coefficient of variation

Criterion for splitting	Mean		Variance		CV	
	r_G	s.e.	r_G	s.e.	r_G	s.e.
Milk yield	0·964	0·071	1·016	0·070	1·109	0·073
Fat yield	0·994	0·068	1·067	0·072	1·125	0·077
Protein yield	0·986	0·078	1·007	0·087	1·075	0·088
Fat content	1·065	0·036	1·032	0·035	1·042	0·035
Protein content	1·017	0·029	1·009	0·027	1·003	0·028
Log_{10} milk yield	0·915	0·067	0·929	0·061	0·982	0·061
Log_{10} fat yield	0·942	0·061	0·959	0·060	0·969	0·062
Log_{10} protein yield	0·938	0·073	0·893	0·075	0·931	0·072

DISCUSSION

Unsplit data. A secondary objective of this study was to obtain genetic parameters for British Friesian adult females (cows) in England and Wales, without regard to level of production, for use in sire evaluation and cow indexing procedures. In particular, estimates of heritability of protein yield and content have not been made on extensive bodies of British data. Although protein content has a higher heritability than fat content (Table 3), protein yield has a slightly lower heritability than fat yield. This is perhaps in part because the genetic correlation between milk yield and protein content is more negative than that between milk yield and fat content, as recent analyses of correlations of predicted breeding values on some of the same data have shown (G. J. T. Swanson, unpublished results). In general, however, the values obtained agree well with published values (Maijala and Hanna, 1974). There is some indication that yields transformed to logarithms (subsequently 'log yields') have a higher heritability than untransformed yields (subsequently just 'yields'), perhaps because of improved homogeneity of phenotypic variances over herds at different production levels (from Table 3). The MMB use untransformed yields in calculations for sire proofs (G. J. T. Swanson, personal communication). The present data

TABLE 5
Summary of estimates of heritability (× 100) of milk yield at different production levels

Source	Population	Low	Medium	High
Maijala and Hanna† (1974)	Various (weighted mean)	21 ± 0·3	26 ± 0·4	28 ± 0·3
Averdunk and Alps‡§ (1971)	Bavarian Fleckvieh	22	25	33
Mokhtar Ibrahim (1979)‡	German Holstein-Friesian	22 ± 2	26 ± 2	22 ± 2
Danell (1981)‡	Swedish Red and White	21 ± 1·0	25 ± 0·9	28 ± 1·3

† Summary of published figures, estimated from both half-sib correlation and daughter-dam regression. They also give mean estimates of 19 ± 3·8 for 'lowest' and 35 ± 1·4 for 'highest' levels.

‡ Estimates from half-sib correlation.

§ No. s.e.'s given. Based on 456 sires, 68 707 daughters.

suggest that use of log yields or, somewhat similarly, the MMB's Production Index, which expresses yields as ratios of herd means (G. J. T. Swanson, personal communication) would increase the accuracy of progeny tests.

There are some deficiencies in the analysis which apply to the whole data set, but which should not influence differences between heritabilities at high and low levels. First, no account was taken of selection among sires of the young bulls, which would reduce the genetic variance among their sons and thus the heritability estimate (Robertson, 1977). In a Best Linear Unbiassed Prediction (BLUP) procedure, where account is taken of such selection (Henderson, 1975), values for unselected populations should strictly be used. There is some indication from British data that these effects of selection on variance were small (G. J. T. Swanson, personal communication). Secondly, during the period of this study, increasing use was being made of Canadian Holstein bulls and their sons in the British Friesian population. In this analysis the young bulls have not been grouped according to the proportion of their genes derived from Holsteins, so some of the genetic variance in yield and composition traits may have come from population differences.

Split by mean level. There have been many previous studies in which herds have been divided by level of production and estimates made of heritabilities at each level and of genetic correlations between levels. Summary values of heritability from the review by Maijala and Hanna (1974) and additional results are given in Table 5. In the majority of studies, and certainly overall, an increase in heritability with higher production levels has been obtained. In the only previous analysis of British data (Robertson *et al.*, 1960) there was, however, no trend of heritability with level, either in their Friesian data (with values of 0·31 ± 0·07, 0·37 ± 0·07 and 0·31 ± 0·07 for low, medium and high levels) with 53 d.f. for sires or in two separate analyses on Ayrshires, having 7 and 10 d.f. (These results are included in Maijala and Hanna's summary figures.) The differences in heritability between high and low levels for log yields in the present analysis, including many more sires but smaller progeny groups than Robertson *et al.* (1960), are statistically significant ($P < 0{\cdot}05$), based on the approximate standard errors. In summary, therefore, there appears to be an increase in heritability of yield of milk, fat and protein at higher production levels in Britain.

There is only a small increase in fat and protein content with increased milk production level (in contrary direction to the well established negative correlations for individual cows or sire progeny groups), and the phenotypic variances are scarcely affected by the partition. In studies where heritability of fat content has been reported at each level it has been essentially the same (Mokhtar Ibrahim, 1979) or shown a small increase in proportion to the values at either level (Mason and Robertson, 1956; Averdunk and Alps, 1971; the present study). Because fat and protein contents have higher heritabilities than yield any such increase is of no practical

consequence because, for example, numbers of daughters required for progeny tests of bulls are determined by the more lowly heritable yield traits.

The analysis of the data within production levels gives us no direct guide as to why the heritability should increase with level, so there has been extensive discussion in the literature (e.g. Mason and Robertson, 1956; Legates, 1962). At its simplest, cows may be allowed to 'express their genetic potential' at the high level (although this phrase has no real interpretation in quantitative genetic models unless environmental deviations are downward), reflecting the recommendations of Hammond (1947). Other, more definable explanations, are better pedigree records (although not attention to pedigree, since the young sires in this analysis had no prior reputation), lower disease incidence and more individual feeding. If the analysis had been conducted in particular environments, for example if feeding systems had been recorded, it might then have been easier to explain the results. For specified environments, studies with other animals e.g. growing mice, have revealed negative associations between performance levels and heritability (e.g. Falconer, 1960).

As in this study, previous estimates of genetic correlations between production at different levels were close to 1·0 or, equivalently, sire × level interactions were found to be small (see Danell, 1981, for a review). In the present data, progeny group sizes were too small to give much power to a test of whether the correlation departed from 1·0 (Robertson, 1959), but some analyses of progeny of 27 widely used sires in the same data set showed negligible interactions (Ahmed, 1981). Overall there seems to be no case for developing strains for different production levels, or indeed for herds classified by variance or coefficient of variation (Table 4).

Split by variance and CV. It is well known that the coefficient of variation of milk production (indeed of many traits) is little affected by changes in mean performance. Thus at high levels of production, variability is also increased, as shown in this analysis (Table 3). It seemed relevant to ask, therefore, whether the increased heritability with yield level could be attributed to the increase in variance, and this study was extended from the traditional classification of herds by mean yield to a classification by variability of yield among cows within seasons, ignoring sires. Preliminary studies had shown that, although mean and variance (or standard deviation) were correlated, there were substantial differences among herds in variance which could not be accounted for by mean level. This is illustrated in Table 2, where it is shown that only two-thirds of herds were high-high or low-low for mean and variance. To effect a more complete separation of mean and variance effects, herds were also split by coefficient of variation, with rather less than half the herds falling in the high-high or low-low category for mean and CV, such that the mean performance of high CV herds was less than that for low CV herds.

The results of these further analyses are rather puzzling. Consider log yields: on the split by variance, within sire variance increased from low to high groups, in contrast to the split on mean, where there was no change; but the between sire variances roughly doubled, such that the heritabilities at each level and increase in heritability (0·095 for log milk yield) were almost exactly the same as for the split on mean. Except for fat and protein contents there was the same, but smaller trend, with the split on CV. Because the heritability was higher, the increased variability in high variance herds cannot be due to environmental errors associated with erratic management or disease, so perhaps the increased variability (the greater part of which is within sires) allows fuller expression of genetic merit. On balance, it appears as though both higher mean and higher variance confer increased heritabilities, with the split on CV indicating that variance is rather more important. Again, however, since herds of different types are being classified solely by statistics rather than management systems, we cannot identify what kind of management system best allows expression of genetic variability, or whether, for example, the effects are due solely to elimination of records of less than 200 days. Further investigations of the causes and consequences of the differences in variance are required.

Practical implications. Because the results give no indication of genotype × environment interaction, we shall only consider the implications for evaluation of breeding value of sires and dams in the national herd.

If estimates of merit of sires used in artificial insemination are made by combining data from all herds without regard to variance, i.e. using a model for BLUP or associated index procedures in which variance is assumed homogeneous, the weight given to different herds is, in effect, proportional to the phenotypic standard deviation within herds. Use of log yields (or ratios of yields to herd mean) equalize the weights at different mean levels of production providing (as Table 3 indicates) the coefficient of variation of yield is independent of mean production. There are, however, residual differences in variance so use of this transformation still gives more weight to 'high variance herds'. What are the optimal weights?

The áccuracy (r) of a progeny test with n effective daughters is $r = (nh^2/[4 + (n-1)h^2])^{1/2}$, so for a test of equal accuracy at high (H) and low (L) levels (however levels are defined),

$$\frac{n_L}{n_H} = \frac{\sigma^2_{BH}/\sigma^2_{WH}}{\sigma^2_{BL}/\sigma^2_{WL}} = \frac{h^2_H/(4-h^2_H)}{h^2_L/(4-h^2_L)} \simeq \frac{h^2_H}{h^2_L} \qquad (4)$$

For herds classified by mean or variance this ratio (4) is approximately 1·2 for yield and 1·4 for log yield (Table 3), so, using log yield, a test with 35 effective daughters at high level is equivalent to a test with 50 effective daughters at low level. There would, therefore, be some benefit in concentrating resources at higher level.

With testing at all levels, straightforward index calculations show that the optimal weight to give to a mean, $\bar{x}$, at some level is $n\sigma_B/\sigma^2_W$, e.g. for two levels,

$$I = \sigma_{BL} n_L \bar{x}_L/\sigma^2_{WL} + \sigma_{BH} n_H \bar{x}_H/\sigma^2_{WH}$$

For these data, the ratio $(\sigma_{BH}/\sigma^2_{WH})/(\sigma_{BL}/\sigma^2_{WL})$ is as follows (from Table 3)

Split by	Mean	Variance	CV
milk yield	0·94	0·86	0·92
log milk yield	1·20	0·98	0·89

This implies that when untransformed data are used slightly lower weights should be given in high mean or high variance herds; or equivalently, when no scaling of data is practised, rather too much weight is given to high variance herds. If log yields are used, it turns out that because the ratio of within sire standard deviation (σ_W) to between sire variance ($\sigma_B{}^2$) is almost the same at both levels of herd variance, the proper weighting is obtained by simple averaging. The same arguments apply to BLUP as to these simple index calculations.

Comparisons among animals from different herds have also to be made when constructing indices for cows, for example when selecting potential bull dams. For simplicity, consider cows in different herds sired by the same bull with an accurate proof. Because three-quarters of the additive genetic variance is within families, the within-family regressions of breeding value on deviation from family mean are $3\sigma^2_{BH}/\sigma^2_{WH}$ and $3\sigma^2_{BL}/\sigma^2_{WL}$ at high and low levels (for some criterion of splitting), with the latter becoming $3(\sigma^2_{BL}/\sigma^2_{WL})/(\sigma_{BH}/\sigma_{WH})$ if standardized to make predictions at high level. The ratio of the two regressions is again $(\sigma_{BH}/\sigma^2_{WH})/(\sigma_{BL}/\sigma^2_{WL})$, so for log yields comparisons between cows can be made without further scaling because this ratio is near 1·0 in the present data. With untransformed data, the better cows in high variance herds are overrated. Even so, when cows are selected on an index, a higher proportion should come from high variance herds.

It must be emphasized, however, that only two levels were considered in the present analysis and the relationship of between and within family variances may not apply over the whole range. Also, all the estimates are subject to sampling errors and the inferences should be treated with caution.

ACKNOWLEDGEMENTS

We are indebted to the Milk Marketing Board of England and Wales for providing the data, to Mr Gordon Swanson for help and advice, and to the Agricultural Research Council for financial support.

REFERENCES

AHMED, M.-K. A. 1981. Genotype × environment interactions in dairy cattle. *Ph.D. Thesis, Univ. Edinburgh.*

AVERDUNK, G. and ALPS, H. 1971. [Interaction between herd level and sire for milk yield of the Fleckvieh breed in Bavaria.] *Z. Tierzücht. ZüchtBiol.* **88:** 197-202.

DANELL, B. 1981. Interaction between genotype and environment in sire evaluation for milk production. *Acta Agric. scand.* **32:** 33-46.

FALCONER, D. S. 1960. Selection of mice for growth on high and low planes of nutrition. *Genet. Res.* **1:** 91-113.

GUIARD, V. and HERRENDÖRFER, G. 1977. Estimation of the genetic correlation coefficient by half sib analysis if the characters are measured on different offsprings. *Biometr. J.* **19:** 31-36.

HAMMOND, J. 1947. Animal breeding in relation to nutrition and environmental conditions. *Biol. Rev.* **22:** 195-213.

HARVEY, W. R. 1977. Users guide for LSML76. Mixed model least-squares and maximum likelihood computer program. Ohio State Univ., Columbus. (Mimeograph).

HENDERSON, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31:** 423-447.

LEGATES, J. E. 1962. Heritability of fat yields in herds with different production levels. *J. Dairy Sci.* **45:** 990-993.

MAIJALA, K. and HANNA, M. 1974. Reliable phenotypic and genetic parameters in dairy cattle. *Proc. Ist Wld Congr. Genet. Appl. Livest. Prod., Madrid, Session 1*, pp. 541-563.

MASON, I. L. and ROBERTSON, A. 1956. The progeny testing of dairy bulls at different levels of production. *J. agric. Sci., Camb.* **47:** 367-375.

MOKHTAR IBRAHIM, S. A. 1979. [Interactions between environmental intensity and genetic potential of Holstein-Friesian crosses in the Friesian population of Lower Saxony.] *Doctoral Thesis, Univ. Gottingen.*

NELDER, J. A. and WEDDERBURN, R. W. M. 1972. Generalized linear models. *Jl R. statist. Soc.* **135(A):** 370-384.

ROBERTSON, A. 1959. The sampling variance of the genetic correlation coefficient. *Biometrics* **15:** 469-485.

ROBERTSON, A. 1977. The effect of selection on the estimation of genetic parameters. *Z. Tierzücht. ZüchtBiol.* **94:** 131-135.

ROBERTSON, A., O'CONNOR, L. K. and EDWARDS, J. 1960. Progeny testing dairy bulls at different management levels. *Anim. Prod.* **2:** 141-152.

(*Received* 27 *March* 1982—*Accepted* 6 *July* 1982)

39

Dairy sire evaluation using a "rolling months" model

by

V.P.S. Chauhan and R. Thompson

Sonderdruck aus Zeitschrift für Tierzüchtung und Züchtungsbiologie
Bd. 103 (1986), H. 5, S. 321–333
VERLAG PAUL PAREY · HAMBURG 1 · SPITALERSTRASSE 12
Rechte, auch die der Übersetzung, des Nachdrucks, der photomechanischen Wiedergabe und der Speicherung in Datenverarbeitungsanlagen, vorbehalten, © 1986 Verlag Paul Parey, Hamburg und Berlin

Institute of Animal Genetics, University of Edinburgh; Animal Breeding Research Organisation, Edinburgh, Scotland

Dairy sire evaluation using a "rolling months" model

By V. P. S. CHAUHAN and R. THOMPSON

Ms. received 15. 12. 1985

Introduction

In most herd-year-season models of sire evaluation, seasons are formed by grouping several consecutive months together. In these models with fixed herd-year effects, it is assumed that all cows calving in different months but the same season have equal covariances with each other, and that the covariances of any of them with any of those calving in a different season are zero. However, in a model in which year effects within herd are random, the covariances between cows calving in the same as well as different seasons are assumed to be non-zero. CHAUHAN (1987b) examined herd-year-season models with herd-year effects as fixed, and season within herd effects as random.

However, since the intra-herd environmental variations from month to month (i.e. the herd-year × month interactions) are usually large (CHAUHAN and HILL 1986) it does not seem to be appropriate to assume that the covariances between all cows calving in different months of a season are equal or are zero. The environmental correlations among milk production records in the same herd-year-season have been found to be slightly smaller when seasons are taken to be longer (CHAUHAN 1987a), showing that the environmental covariance between cows decreases as calving dates become further apart. Van Vleck (1966) reported the environmental correlations between milk yields of cows calving from zero to 18 year-seasons apart. No specific trends in the correlations were observed as the time between year-seasons increased. However, his estimates were biased because they involved the covariances of each record with the rest in a herd, since each record was expressed as a deviation from the herd mean.

This study was undertaken: (1) to estimate the covariances between the first lactation records of cows calving in the same month and between records of those calving various months apart, and (2) to examine the usefulness of a sire evaluation model accounting for the variances and covariances of months within herd-years.

Material and methods

Data

The same eight subsets of data as described by CHAUHAN (1987b) were used for estimating sire effects. These subsets were generated from a large data set which was split at random by herd number such that all records of a herd were included in any one of the subsets. Data were the first lactation fat yield records, precorrected for the effect of month of calving, of 49242 progeny of 69 widely used proven Holstein-Friesian sires in 1628 herds participating

U.S. Copyright Clearance Center Code Statement: 0044–3581/86/0305–0321 $ 2.50/0
J. Anim. Breedg. Genet. 103 (1986) 321–333
© 1986 Verlag Paul Parey, Hamburg und Berlin
ISSN 0044–3581

in the Dairy Progeny Testing Scheme (DPTS) of the Milk Marketing Board of England and Wales (MMB), from November 1972 to October 1981. Each subset contained records of all 69 sires. The data on widely used sires were useful for empirical comparisons of sire effects from different subsets.

The variances and covariances of month effects, however, were estimated from the data between December, 1972 and November, 1980 (43089 records in 1533 herds), in view of the current definition of the recording year – "December–November" – used by the MMB. In order to have an empirical check on the method it was considered more appropriate to estimate variances and covariances of months independently from two sets (22077 and 21012 records), each of which represented 4 subsets, than estimating from all data together.

Estimating variances and covariances of months within herd-year

Each record was represented by the following model:

$$Y_{ijklm} = \mu + hy_{ij} + M_k + (hy \times M)_{ijk} + Z_l + bA + b'A^2 + e_{ijklm}$$

where:

Y_{ijklm}	=	record of the mth daughter of the lth sire calved at age A in the jth year and the kth month of the ith herd,
μ	=	overall mean,
hy_{ij}	=	random joint effect of the ith herd and the jth year of calving,
M_k	=	fixed effect of the kth month of calving,
$(hy \times M)_{ijk}$	=	random effect common to records of cows calved in the kth month of the jth year of the ith herd,
Z_l	=	fixed effect of the lth sire (fixed effect since proven sires),
b and b′	=	linear and quadratic regression coefficients on age at calving, A,
e_{ijklm}	=	random error.

This model was fitted to the data using the LSML76 computer program of Harvey (1977). Using the least squares constants from the above analysis, the data were adjusted for the effects of sire, month and age at calving. The herd-year effects were removed by subtracting the respective herd-year means (calculated from records adjusted for sire, month and age) from each record. The herd-year mean included the record itself also. These deviated records were then used to calculate sums of squares (SSQ) for each month and sums of crossproducts (SCP) between different records in the same month and between different months. All calculations of SSQ and SCP were done on a herd-year basis, therefore the SSQ and SCP and their coefficients (the numbers on which these SSQ and SCP were based) were accumulated, accordingly. Altogether, there were 12 SSQ and 12 SCP for the same month (i.e. 0-month-apart), and 66 SCP for different months. The value 66 consists of 11 1-month-apart SCP and 10, 9, 8, 7, 6, 5, 4, 3, 2 and 1, respectively, 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10- and 11-months-apart SCP, the 11-months-apart SCP being only between December and November. Since the herd-year mean are subtracted from each record, these SSQ and SCP include covariances with other months, in addition to covariance between the months in question. Therefore, the equations of the expectation of these SSQ and SCP, in terms of the unknown variance and covariances, and the procedure for solving them are given in the appendix.

It was assumed that the covariances of records between any pair of months the same time apart are equal. For example, taking the recording year "December–November" currently used by the MMB, the one month apart covariance is the covariance of records of December with January, January with February, ..., October with November; and the two months apart: December with February, January with March, ..., September with November, and so on.

The estimates of 0 to 11 months apart covariances obtained from both sets are plotted in

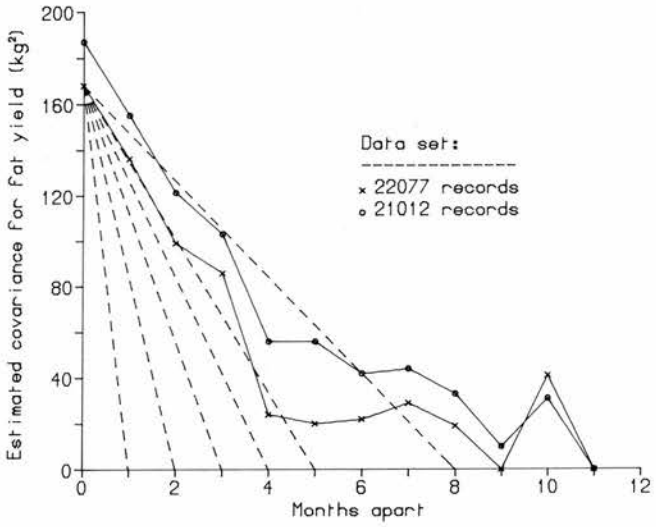


Fig. 1. The trends in the various months apart observed covariances, and alternative trends investigated; ——— observed, – – – – investigated

Figure 1. The trends in the estimates of covariances from both sets are quite similar. The covariances from 0 through to 5-months-apart showed an almost linear decline, and thereafter all covariances were apparently similar and close to zero. The variation in the estimates of 6 to 10-months-apart covariances (since all the covariances are not estimable, the 11-months-apart covariance was constrained to be zero) can be attributed to the fact that, as compared to the 0- to 5-months-apart covariances, they were based on smaller coefficients, e.g. the 10-months-apart covariance comprised the covariance between records in December and October, and January and November.

Using the same data CHAUHAN (1985) estimated the various months apart covariances from the analyses over 24 and 96 months. Those estimates also showed a linear decline from 0 through to 5 months apart. All covariances beyond 11 months apart showed a gradual declining trend, although very slow, suggesting periodic trends over years. The estimates of covariances from these two analyses were biased upward as they included the covariances between years within herd-period (period being of 24 and 96 months).

Sire evaluation

The observed covariances between months could have been taken into account in sire evaluation by BLUP using a procedure similar to the sires's relationships matrix (HENDERSON 1975a). But in this study it was decided to derive a model in which it is assumed that the covariances between months decline linearly as the distance between them increases. An assumption that the covariances between months decrease linearly from 0 through to 5 months apart and all covariances beyond 5 months apart are zero is close to the observed relationship between the covariance and the time between-months (Figure 1). The above assumptions would be helpful from the point of view of computations as the covariances between several pairs of months can be taken to be zero, a priori. However, in order to investigate the appropriate model alternative analyses were conducted assuming that the covariances between-months decline linearly from 0 through to either 1- or 2- or 3- or 4- or 5- or 8-months-apart and the rest of the between-months covariances are zero. Models such as this have been suggested by BARTLETT (1978) and WILKINSON, ECKERT, HANCOCK and

MAYO (1983) for analyses of crop variety trials. This model hereafter will be called the "rolling months" model as it accounts for the covariances of each month with a "specified" number of months on each side of it. Suppose the covariances between-months are assumed to decline linearly from 0 through to W months apart, then the covariances among '2W-1' months are accounted for with each individual month – with the exception of those of the beginning or of the end of the herd-year blocks. The value of W will hereafter be referred to as the "width" of a group of months among which the covariances are accounted for. In order to obtain the assumed variance-covariance structure between months a "pseudo-month" (P) effect was fitted. Then the month effects (m) in terms of the pseudo-month effects can be given by the following equations,

$$m_i = P_i + P_{i+1} + \ldots + P_{i+W-1}$$
$$m_j = P_j + P_{j+1} + \ldots + P_{j+W-1}$$

with variance of month effects equal to $W\Phi$ (Φ being the variance of records in the same pseudo-month), and covariance of month effects in the same herd-year i-months-apart equal to $(W - i)\Phi$ for $1 \leqslant i < W$ and equal to 0 for $i \geqslant W$. The following assumptions are made in fitting the rolling months model:

(i) homogeneity of variance over pseudo-months ($\mathrm{Var}[P] = \Phi$)

(ii) $\mathrm{Cov}(P_i, P_j) = 0; i \neq j$

Using the above assumptions the linear relationship between the covariance (Cov) and different distances apart (shown in Figure 1) can be demonstrated as follows, using an example with information over only 4 months and assuming W equal to 3,

$$\mathrm{Cov}(m_1, m_1) = C_0 = \mathrm{Cov}(P_1 + P_2 + P_3, P_1 + P_2 + P_3) = 3\Phi$$
$$\mathrm{Cov}(m_1, m_2) = C_1 = \mathrm{Cov}(P_1 + P_2 + P_3, P_2 + P_3 + P_4) = 2\Phi$$
$$\mathrm{Cov}(m_1, m_3) = C_2 = \mathrm{Cov}(P_1 + P_2 + P_3, P_3 + P_4 + P_5) = \Phi$$
$$\mathrm{Cov}(m_1, m_4) = C_3 = \mathrm{Cov}(P_1 + P_2 + P_3, P_4 + P_5 + P_6) = 0$$

where:

$$C_i = \text{i-months-apart covariance}$$

The records were assumed to be represented by the following model for sire evaluation using rolling months.

$$y = Pb_1 + Hb_2 + Cb_3 + Zu + e \qquad (1)$$

where,

- y = vector of the first lactation fat yield records,
- P = design matrix for pseudo-months resulting from the assumed variance-covariance structure between months,
- H = design matrix for herd-years,
- C = matrix of information on covariables (i.e. age and square of age at calving),
- Z = design matrix for sires,
- b_1 = vector of unknown random effects for pseudo-months,
- b_2 = vector of unknown fixed effects for herd-years,
- b_3 = vector of unknown linear and quadratic regression coefficients of fat yield on age at calving,
- u = vector of unknown fixed effects for sires, (fixed since proven sires),
- λ = ratio of residual variance to the variance of records within the same pseudo-month, σ_e^2/Φ
- e = vector of residuals (random)

The rolling months model has been compared with a model examined by CHAUHAN (1987b) in which herd-year effects were regarded as fixed and herd-year-month effects as random

and uncorrelated. This model will be referred to as the "random month" model, hereafter. A rolling months model, in which the covariances among records only in the same month are accounted for and the rest of the covariances between months are assumed to be zero (i.e. W = 1), is equivalent to the random month model.

The mixed model equations (MME) pertaining to the rolling months model are given below in matrix notation:

$$\begin{bmatrix} P'P + \lambda I & P'H & P'C & P'Z \\ H'P & H'H & H'C & H'Z \\ C'P & C'H & C'C & C'Z \\ Z'Z & Z'P & Z'C & Z'Z \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ u \end{bmatrix} = \begin{bmatrix} P'y \\ H'y \\ C'y \\ Z'y \end{bmatrix} \qquad (2)$$

All the submatrices in (2), except those in the equations for pseudo-months, can be constructed by the usual procedure. The total number of pseudo-month effects to be fitted within herd-year will be 'm + W − 1' where 'm' is the number of months. Assuming that the data are available over only 4 months in a herd-year and the assumed value of W is equal to 3, an example is given below to construct the matrices included in the mixed model equations for estimating the pseudo-month effects (i.e. P'P, P'H, P'C, P'Z, P'y matrices in equation [2]). Suppose n_1, n_2, n_3 and n_4 are the numbers of records respectively for months m_1, m_2, m_3 and m_4, then the P'P and P'H matrices will be as follows. The submatrices P'C, P'Z and P'y can also be constructed similarly.

Pseudo-month

$$P'P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} n_1 + \lambda & n_1 & n_1 & 0 & 0 & 0 \\ n_1 & n_1 + n_2 + \lambda & n_1 + n_2 & n_2 & 0 & 0 \\ n_1 & n_1 + n_2 & n_1 + n_2 + n_3 + \lambda & n_2 + n_3 & n_3 & 0 \\ 0 & n_2 & n_2 + n_3 & n_2 + n_3 + n_4 + \lambda & n_3 + n_4 & n_4 \\ 0 & 0 & n_3 & n_3 + n_4 & n_3 + n_4 + \lambda & n_4 \\ 0 & 0 & 0 & n_4 & n_4 & n_4 + \lambda \end{bmatrix} \end{array}$$

$$P'H = \begin{bmatrix} n_1 \\ n_1 + n_2 \\ n_1 + n_2 + n_3 \\ n_2 + n_3 + n_4 \\ n_3 + n_4 \\ n_4 \end{bmatrix}$$

Since the submatrix P'P + λI is not a diagonal matrix the equations for the pseudo-month effects may be absorbed either by inverting it or by successively eliminating the pseudo-month effects. In this study the latter procedure was used as it required less computing time than the inversion. The MME in (2) after absorption of the equations for the effects of pseudo-month and herd-year can be written as below,

let:

$$S = [I - P(P'P + \lambda I)^{-1} P']$$
$$Q = [S - SH(H'SH)^{-1} H'S]$$

then the equations for age and sire effects would be,

$$\begin{bmatrix} C'QC & C'QZ \\ Z'QC & Z'QZ \end{bmatrix} \begin{bmatrix} b_3 \\ u \end{bmatrix} = \begin{bmatrix} C'Qy \\ Z'Qy \end{bmatrix} \qquad (3)$$

The estimates of sire effects were obtained by absorbing the equations for age into sire equations and then inverting the sire's coefficient matrix,

letting: $$R = [Q - QC(C'QC)^{-1} C'Q]$$

then

$$u = (Z'RZ)^{-1}(Z'Ry)$$

The residual mean squares (σ_e^2) were calculated by the procedure given by Thompson (1969).

$$\sigma_e^2 = [\sum_{i=1}^{h} y'Qy - y'QC(C'QC)^{-1}C'Qy - y'RZ(Z'RZ)^{-1}Z'Ry]/[N - \text{rank (fixed effects)}] \quad (4)$$

where, h = total number of herd-years
N = total number of records

$$V(u) = (Z'RZ)^{-1}\sigma_e^2 \quad (5)$$

The MME for the random month model would be similar to (2) and (3) of the rolling months model. These two models differ only in assuming the variances and covariances between months. In the random month model all covariances between months are assumed to be zero, consequently, the $P'P + \lambda I$ submatrix in (2) for the month equations in the random month model will be diagonal. Therefore, in this study only these two models will be compared with each other.

Comparison of rolling months and random month models

The following two criteria were used to compare the rolling months model with the random month model:

(a) Empirical variance: the following formula was used to calculate the empirical variances for each sire.

$$\text{Empirical variance} = [\sum_{j=1}^{P} (u_{ij} - \bar{u}_i)^2]/(P - 1) \quad (6)$$

where:

$$\bar{u}_i = (\sum_{j=1}^{P} u_{ij})/P$$

u_{ij} = sire effect for the ith sire from the jth subset of data,
P = number of subsets of data

(b) Comparing the predicted variances of the estimates of sire effects given that the rolling months model is the true model (see HENDERSON 1975b). Let the random month model be model 1 and the rolling months model be model 2. Then the variance-covariance (V) of elements in the vector y in (1) for both models can be written as follows:

$V_1 = I\sigma_e^2 + MM'\sigma_m^2$ ... (random month model)
$V_2 = I\sigma_e^2 + PP'\Phi$ ... (rolling months model)
where: M = design matrix for herd-year-months
σ_m^2 = variance of records in the same month

Let the variance-covariance matrices of month effects for model 1 and 2 be denoted by A_1 and A_2, respectively.
Then

$$V_1 = I + MA_1M'$$
$$V_2 = I + MA_2M'$$

and

$$V_2 - V_1 = M(A_2 - A_1)M'$$

Based on the same design matrix the predicted variances of the estimates of sire effects on the random month (model 1) and the rolling months (model 2) models can be written as, let:

$$(X'V^{-1}X) = (Z'RZ)$$

then:

$$\text{Var}(u_1 \mid \text{model}_1) = (X'V_1^{-1}X)^{-1}\sigma_{e1}^2 \quad (7)$$
$$\text{Var}(u_2 \mid \text{model}_2) = (X'V_2^{-1}X)^{-1}\sigma_{e2}^2 \quad (8)$$
$$\text{Var}(u_1 \mid \text{model}_2) = [(X'V_1^{-1}X)^{-1}X'V_1^{-1}V_2V_1^{-1}X(X'V_1^{-1}X)^{-1}]\sigma_{e2}^2 \quad (9)$$

When $V_2 = V_1$, (9) reduces to (7). Given that model 2 (the rolling months model) is the true model, the efficiency of model 1 (the random month model) can be judged by comparing $\text{Var}(u_2 \mid \text{model}_2)$ and $\text{Var}(u_1 \mid \text{model}_2)$. The larger the value of $\text{Var}(u_1 \mid \text{model}_2)$ in relation to $\text{Var}(u_2 \mid \text{model}_2)$ the more efficient the rolling months model.

Computing procedure for $\text{Var}(u_1 \mid \text{model}_2)$

Equation (9) can be rewritten as follows:

$$\text{Var}(u_1 \mid \text{model}_2) = [(X'V_1^{-1}X)^{-1} + (X'V_1^{-1}X)^{-1}X'V_1^{-1}(V_2-V_1)V_1^{-1}X(X'V_1^{-1}X)^{-1}]\sigma_{e2}^2 \quad (10)$$

Replacing V_2-V_1 by $M(A_2-A_1)M'$:

$$\text{Var}(u_1 \mid \text{model}_2) = [(X'V_1^{-1}X)^{-1} + (X'V_1^{-1}X)^{-1}(X'V_1^{-1}M(A_2-A_1)M'V_1^{-1}X(X'V_1^{-1}X)^{-1}]\sigma_{e2}^2 \quad (11)$$

In order to compute the quantities in (11) the coefficient matrix in (3) for the random month model, given that the rolling months model is the true model, would be of the form given below.
Let:

$S = [I - M(M'M+\lambda I)^{-1}M']$ ... (adjustment for random effect of herd-year-month)
$Q = [S - SH(H'SH)^{-1}H'S]$ ... (adjustment for fixed effect of herd-year)

then the coefficient matrix will be as follows:

$$\begin{bmatrix} C'QM(A_2-A_1)M'QC & C'QM(A_2-A_1)M'QZ \\ Z'QM(A_2-A_1)M'QC & Z'QM(A_2-A_1)M'QZ \end{bmatrix}$$

where:

M = design matrix for herd-year-months
$\lambda = \sigma_e^2/\sigma_m^2$ on the random month model

The implications of the rolling months models for different herd sizes were also investigated, using herd sizes defined as being either large, moderate or small. The data subsets with the "moderate" herd sizes were simply the data subsets used in the previous parts of this study. To simulate the situations of large herd sizes all herd-year subclasses having less than 8 records were discarded from the initial data subsets. To simulate the situations of small herd sizes each subset was split at random into two, by animal number. This resulted in a total of 16 subsets.

Results and discussion

The average values of the percent increases in the empirical variances for each sire on the random month model over the rolling months model showed that the rolling months models

Table 1. **Percent increases in the empirical, and predicted variances of the estimates of sire effects on the random month model given that the rolling months models with different widths** (W) are the true models.

Subset	W = 1 [1]λ = 4.12	W = 2 λ = 8.24	W = 3 λ = 12.36	W = 4 λ = 16.48	W = 5 λ = 20.60	W = 5 [2]λ = 50.00	W = 8 λ = 32.95
			% increase in predicted variances				
1	0.75	1.25	1.63	1.92	2.05	0.74	2.11
2	0.86	1.48	1.93	2.25	2.42	0.87	2.32
3	0.82	1.36	1.89	2.23	2.46	0.89	2.54
4	0.76	1.27	1.60	1.85	1.99	0.76	2.05
5	0.69	1.07	1.41	1.60	1.77	0.68	1.76
6	0.76	1.27	1.65	1.87	2.05	0.71	1.95
7	0.74	1.26	1.79	2.12	2.33	0.75	2.36
8	0.78	1.42	1.92	2.23	2.46	0.89	2.58
Average	0.77	1.30	1.73	2.01	2.19	0.79	2.21
			% increase in the empirical variances[3]				
	2.29	3.19	3.47	3.78	3.95	3.40	4.29

[1] $\lambda = \sigma_e^2/\Phi$ added to the diagonal of the P'P submatrix in rolling months model. The value of λ as the ratio of residual variance to variance of records in the same month (σ_e^2/σ_m^2) used in the random month model was 8.07. The σ_e^2 and the covariance of records in the same month (C_0) respectively were 690 and 168 kg^2.

[2] The value of λ equal to 50.00 is an arbitrary value.

[3] Average values of the % increase in empirical variances for random month model over those for the rolling months model.

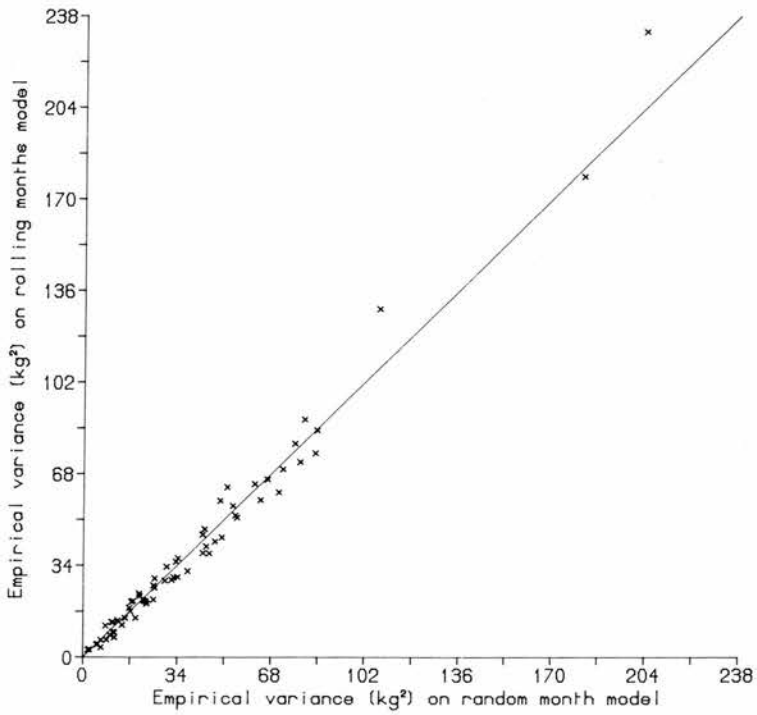


Fig. 2. Empirical variances for each sire plotted for the rolling months model (W = 5 months) against those for the random month model. The straight line has a slope of 1.00

with all values of W (i.e. the width for rolling) examined were more efficient than the random month model (Table 1). As the value of W increased the rolling months model became slightly more efficient. The empirical variances for each sire for the rolling months model (W = 5) have been plotted against those for the random month model in Figure 2. The straight line drawn through the origin has a slope of 1.00. The number of sires having empirical variances below the straight line is more than those above the line, showing that the empirical variances on the rolling months model were smaller. The plots of empirical variances with different values of W have been given by CHAUHAN (1985). All plots were consistent with the average values of the percent increases in the empirical variances given in Table 1. The values of empirical variances using an arbitrary value of λ equal to 50 and W equal to 5 are also given in Table 1. Substantial changes in the value of λ over the range 10 to 70 had only a marginal effect on the empirical variances.

Comparisons of the rolling months models (fitting different widths) with the random month model in terms of the predicted error variances of the estimates of sire effects using (8) and (9) are also given in Table 1. Average increases in the predicted variances of the estimates of sire effects on the random month model, given that the rolling months models with a width either 1, 2, 3, 4, 5 or 8 months were the true models, were estimated to be respectively 0.77, 1.30, 1.73, 2.01, 2.19 and 2.21%. These values can be interpreted as the percent gain in the efficiency of using the rolling months model over the random month model. The estimates of predicted and empirical variances were fairly consistent with each other. Although advantages were small, the rolling months model became more efficient as the width of the rolling months increased. The advantages of using a rolling months model with a width of 8 months over a width of 5 months were not as much as were observed for the increase in the widths from 1 to 5 months. Therefore a rolling months model with a width equal to 5 months seems to be optimal. The rolling months model with a value of λ equal to 50 (i.e. an arbitrary value 2.5 times higher than the estimated value of λ with W = 5) became closer to the random month model in terms of predicted as well as empirical variances. Its efficiency, in terms of the predicted variances, over the random month model reduced from 2.19% to 0.79%.

Table 2. **Percent increases in the empirical, and predicted variances of the estimates of sire effects on the random month model given that the rolling months model with W = 5 and λ = 20.60 is the true model, for all situations with different herd sizes.**

[1]Herd size	No. of subsets	Average herd-year size	% increase in the empirical variances on the random month model over the rolling months model	% increase in predicted variance on random month model	[2]Product moment correlation
Small	16	4.4	1.32	1.42	0.997
Moderate	8	7.8	3.95	2.19	0.994
Large	8	14.8	1.33	2.55	0.994

[1] The data on widely used proven sires are regarded as being from herds of moderate size. Small herds were created by splitting each subset at random into two, and large herds by discarding all herd-year subclasses having less than 8 records.

[2] Product moment correlation between the estimates of sire effects on the rolling months and the random month models.

The results of the investigation on the usefulness of the rolling months model over the random month model for different herd size situations are given in Table 2 and Figure 3. The empirical variances on the random month model were larger than those for the rolling months model for all herd size situations. For the predicted variances it was observed that the rolling months model was more efficient than the random model, for larger herds. Figure 3 shows the percentage increases in the predicted variances of the estimates of sire effects, from

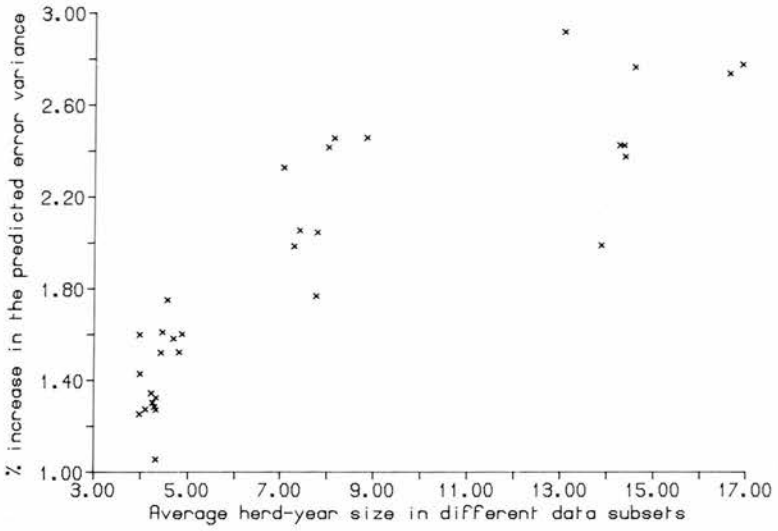


Fig. 3. Percent increase in the predicted error variance of sire effects on the random month model given that rolling months model (W = 5 months) is the true model for all herd size situations. The ratio of residual variance to the variance of records in the same pseudo-month used for these analyses was 20.60

each subset having different herd sizes, on the random month model given that the rolling months model is the true model. The gains in the efficiency of the rolling months model increase slightly as the herd size increases, however after a herd-year size of about 10 the gains appear to asymptote. The average values of the observed gains in the efficiency of the rolling months model over the random month model are 1.42, 2.19 and 2.55%, respectively, for average herd-year sizes of 4.4, 7.8 and 14.8 records (Table 2). The average herd-year size of 14.8 in the data subsets used for large herds in these analyses is incidently equal to the actual number of heifer records available for sire analyses per herd per year in the DPTS herds of the MMB (MMB, 1984). Since the product moment correlations between the estimates of sire effects from both models for all herd sizes were very close to unity (Table 2), however, they would give essentially the same ranking of sires.

It was concluded that the rolling months procedure did not show as much improvement over the random month model as was originally expected. However, the observed and predicted values were consistent. It was observed that there were only 26% of the total number of herd-year-month subclasses filled (in the data subsets for herds of moderate size i.e. 7.8 heifer records per herd per year), and therefore there were not enough connections between months within herd-year to gain the advantages from the variance-covariance structure assumed in the rolling months model. For larger herds there were better connections between months and this was the main reason why the rolling months model was slightly more efficient than the random month model. These results suggest that, compared to the random month model, a model assuming covariances between months would not be advantageous for dairy sires analyses unless herds were large. However the random month model by itself was found to be 37.3% more efficient in terms of empirical variances, and 16% in terms of predicted variances of sire effects than the herd-year-season fixed effects model with each season comprising 4 months (CHAUHAN 1987b).

Appendix

The data are assumed to be precorrected for all effects in the model except those for herd-year effects. Corrections for herd-year effects are made by subtracting the respective herd-year means from each

record. Then the sum of squares (SSQ) and sum of cross-products (SCP) between and within months have expectations as follows:

For the same month:

$$E(SSQ) = n_i E(Y_{ij} - Y..)^2$$
$$= n_i V - 2n_i E(Y_{i.},Y..) + n_i E(Y..)^2 \quad [1]$$

$$E(SCP) = n_i(n_i - 1)C_0 - 2n_i(n_i - 1)E(Y_{i.},Y..) + n_i(n_i - 1)E(Y..)^2 \quad [2]$$

For different months:

$$E(SCP) = E[(n_i n_j)(Y_{i.} - Y..)(Y_{j.} - Y..)] \quad [3]$$
$$= n_i n_j[E(Y_{i.},Y_{j.}) - E(Y_{i.},Y..) - E(Y_{j.},Y..) + E(Y..)^2]$$

where:

$$E(Y_{i.},Y..) = [n_i V + n_i(n_i - 1)C_0 + \sum_{j \neq i} n_i n_j C|i - j|/n_i N$$

$$E(Y_{i.},Y_{j.}) = C|i - j|$$

$$E(Y..)^2 = [\sum_{i=1}^{P} n_i V + \sum_{i=1}^{P} n_i(n_i - 1)C_0 + 2\sum_{i>j}^{P} n_i n_j C|i - j|/N^2$$

E = denotes the sign for expectation
Y_{ij} = the jth record in the ith month adjusted for the effects of sire, month and age of calving
$Y_{i.}$ = mean of the ith month within the herd-year (i = 1, 2, 3, ..., 12)
$Y_{j.}$ = mean of the jth month within the herd-year (j = 1, 2, 3, ..., 12)
Y.. = herd-year mean
N = total number of records in the herd-year
P = total number of months (i.e. 12) within the herd-year
n_i = number of records in the ith month within the herd-year
n_j = number of records in the jth month within the herd-year
C_0 = 0-month-apart covariance i.e. the covariance between records in the same month
V = residual variance within month plus C_0
$C|i - j|$ = the $|i - j|$ months-apart covariance, where $i \neq j$; and the sign | | denotes the absolute value

The equations for the expectations of the 90 terms of SSQ and SCP within a herd-year, in matrix notation, can be written as,

$$Xb = Y \quad [4]$$

Where

X = matrix (order 90 × 13) of the coefficients of unknown parameters of variance and covariances,
b = vector (order 13 × 1) of unknown parameters of variance and covariances, the first parameter being the V (as defined in [1], [2] and [3]) followed by 0 to 11 months apart covariances,
Y = vector (order 90 × 1) of SSQ and SCP based on records deviated from herd-year mean, the first 12 elements are the respective SSQ for different months followed by 11 elements for 0-month-apart SCP, 10 for 1-month-apart, ..., 1 for 11-months-apart SCP.

It was observed that all rows of the matrix X sum to zero but neither the columns of matrix X nor the elements of vector Y sum to zero. According to [1], [2] and [3], altogether there are 90 equations in [4] for the expectations of SSQ and SCP. The matrix X and the vector Y can be transformed to reduce the total number of equations to a number equal to the number of unknown parameters in the vector b. All equations derived from [1] can be added into a single equation and all those from [2] into a second equation. Likewise the 66 equations derived from [3] are reduced to only 11 equations since the equations pertaining to the expectations of the same months apart SCP are added together. These 11 equations need to be multiplied by two in order to make the transformed matrix X (i.e. X^*) symmetric. This transformed matrix X^* is now of order 13 × 13 and elements of each row and each column sum to zero. Similarly, elements of the transformed vector Y^* (order 13 × 1) also sum to zero. The equations pertaining to 1, 2, 3, ..., 11 months apart unknown covariances are multiplied by two since the expectations of the between months SCP of only one triangle of the matrix of SCP were calculated in order to minimize the computations. Then the solution to unknown parameters of variance and covariances in the vector b can be obtained as follows:

$$X^*b = Y^*$$
$$b = (X^*)^{-1}Y^* \quad [6]$$

Because of using deviations within herd-year all covariances can not be estimated i.e. matrix X^* is not of full rank. Therefore, in order to solve these the 11-months-apart covariance was constrained to be zero.

Acknowledgements

The authors are grateful to Prof. W. G. Hill for helpful comments and discussions, and to the Milk Marketing Board of England and Wales, Thames Ditton, Surrey, for providing the data. VPSC was supported by a Commonwealth Scholarship.

Summary

A sire evaluation model assuming covariances among records in the same and different months within herd-year ("rolling months" model) has been compared with a model assuming covariances among records only in the same month and ignoring all covariances between months ("random month" model). The effects of herd-year (fixed), sire (fixed, since the data used were on progeny of proven sires), the overall effects of month of calving (fixed), and age at calving as a covariable by linear and quadratic regressions were the same in both models. The data comprised 8 independent subsets of the first lactation fat yield records on a total of 49242 progeny of 69 widely used proven Holstein-Friesian sires in 1628 herds in England and Wales. The rolling months model, assuming a linear decline in the covariances from 0 through to 5 months apart and the rest of the covariances between months zero, was observed to be only 2.19% more efficient than the random month model. These data had an average herd-year size of 7.8. Analyses for different herd size situations showed that this model was 1.42 and 2.55% more efficient for herds with 4.4 and 14.8 records per year, respectively. However it was concluded that this model would not be advantageous over the random month model for many dairy sires analyses as the average herd-year size in most situations is less than 14.8.

Résumé

Evaluation de la descendance par le modèle »rolling months«

Un modèle d'évaluation de taureaux, admettant les covariances entre résultats des mêmes et différents mois en-dedans du herd-year (modèle »rolling months«) a été comparé avec un modèle admettant les covariances entre résultats uniquement dans le même mois et en ignorant toutes les covariances entre les mois (modèle »random month«). Les effects du herd-year (fixed), taureau (fixed, puisque les dates utilisées provenaient de taureaux testés), les effets du mois de vêlage (fixed) et l'âge au vêlage, une cavariable par regressions linéaire et carrée, étaient identiques dans les deux modèles. Le matériel comprenait 8 présentations indépendantes des rendements en graisse de la première lactation d'un total de 49242 descendants de 69 taureaux testés Holstein-Friesian très utilisés en 1628 troupeaux en Angleterre et au pays de Galles. Le modèle »rolling months«, admettant un déclin linéaire des covariances entre vêlage, séparés de 0 à 5 mois, et une covariance de 0 entre les mois, était uniquement 2,19% plus efficient que le modèle »random month«. Ces dates avaient une grandeur moyenne du herd-year de 7,8. Des analyses de différentes situations du volume des troupeaux montraient que ce modèle était 1,42 et 2,55% plus efficient pour des troupeaux avec 4,4 et 14,8 résultats par an. On en conclut, que ce modèle ne sera pas avantageux sur le modèle »random month« pour beaucoup d'épreuves de la descendance, puisque la grandeur moyenne du herd-year est, dans la plupart des situations, inférieure à 14,8.

Resumen

Evaluación de padres en ganado lechero usando um módelo »rotatorio de meses« (rolling months model)

Un módelo en la evaluación de padres asumiendo las covarianzas entre los registros en los mismos y diferentes meses dentro de rebaño-año (rolling months model), han sido comparados con un módelo que asume las covarianzas entre los registros del mismo mes e ignorando todas las covarianzas entre meses (módelo con meses al azar). En ambos módelos fueron usados los mismos efectos: rebaño-año (fijo), padre (fijo, debido a que los datos provenian de progenies de padres probados), el efecto general del mes de parto (fijo) y la edad al parto como covariables mediante regresion linear y quadrática. Los datos incluyeron 8 grupos independientes, de primeras láctancias registradas de un total de 49242 progenies de 69 padres altamente probados de la raza Holstein-Friesian, en 1628 rebaños en Inglaterra y Wales. El módelo rotatorio de meses, asumiendo una disminución líneal en las covarianzas desde el mes cero hasta el quinto y el resto de las covarianzas entre los meses fue cero, se observó solamente un 2.19% de mayor eficiencia que en el módelo de meses al azar. Estos datos tuvieron un promedio de rebaño-año de 7.8. Un análisis de diferentes tamaños en los rebaños, muestra que este módelo fue 1.42 y 2.55% más

eficiente para los rebaños con 4.4 y 14.8 registros por año, respectivamente. Sin embargo, se puede concluir que este módelo puede ser desventajoso sobre el módelo de meses al azar para muchos análisis de padres en ganado lechero cuando el promedío del tamaño en rebaño-año en muchas situaciones es menor a 14.8.

Zusammenfassung

Stiernachkommenschaftsbewertung mit „Gleitendem Monate"-Modell

Ein Modell zur Bewertung von Stieren, bei dem Kovarianzen zwischen Leistungsabschlüssen in gleichen und verschiedenen Monaten innerhalb eines Herdenjahres („Gleitendes Monate"-Modell) wurde mit einem Modell verglichen, bei welchem Kovarianzen nur zwischen Leistungsabschlüssen des selben Monats und bei Vernachlässigung aller Kovarianzen zwischen Monaten (zufälliges Monatsmodell) angenommen worden sind. Die Wirkungen des Herdenjahres (fixed), Stier (fixed, da die verwendeten Abschlüsse von nachkommengeprüfter Stiere stammten), die Wirkung von Kalbemonaten (fixed) und des Alters beim Abkalben, also einer Kovariable, der durch lineare und quadratische Regression Rechnung getragen wurde, waren in beiden Modellen identisch. Das Material umfaßte 8 unabhängige Aufstellungen von Fettleistungen der ersten Laktation von insgesamt 49242 Nachkommen von 69 stark verwendeten nachkommenschaftsgeprüften Holstein Friesian in 1628 Herden in England und Wales. Das Gleitende Monate-Modell, bei dem eine lineare Verminderung der Kovarianzen zwischen Abkalbungen, die 0 bis 5 Monate separiert waren, und eine Kovarianz von 0 zwischen weiter entfernten Abschlüssen unterstellt wurden, konnte nur 2,19% mehr Varianz erklären als das Zufallsmonatsmodell. In dem Material war die durchschnittliche Herdenjahrkuhzahl 7,8. Analyse verschiedener Herdengrößensituationen zeigte, daß die Varianzminderung des Modells 1,42 und 2,55% bei Herden mit 4,8 und 14,8 Abschlüssen pro Jahr betrug. Als Schlußfolgerung ergab sich, daß das Modell keinen Vorteil über dem Zufallsmonatsmodell bei vielen Nachkommenschaftsprüfungen haben wird, da die durchschnittliche Herdenjahresgröße in den meisten Situationen geringer als 14,8 ist.

References

BARTLETT, M. S., 1978: Nearest neighbour models in the analysis of field experiments. J. Roy. Stat. Soc. B **40,** 147–174.

CHAUHAN, V. P. S., 1985: Environmental adjustments in estimation of breeding value of dairy sires. Ph. D. thesis, Edinburgh University, Scotland.

CHAUHAN, V. P. S., 1987a: Partitioning of herd, year, and season variation in milk production. Livest. Prod. Sci. 16 (in press).

CHAUHAN, V. P. S., 1987b: Dairy sire evaluation fitting some of the herd-year-season effects as random. Livest. Prod. Sci. 16 (in press).

CHAUHAN, V. P. S.; HILL, W. G., 1986: Seasonal grouping in a herd-year-season model of sire evaluation. Anim. Prod. **43,** 63–71.

HARVEY, W. R., 1977: Users guide for LSML76. Mixed model least-squares and maximum likelihood computer program. Ohio State Univ., Columbus. (Mimeograph).

HENDERSON, C. R., 1975a: Use of relationship among sires to increase accuracy of sire evaluation. J. Dairy Sci. **58,** 1731–1738.

HENDERSON, C. R., 1975b: Comparison of alternative sire evaluation methods. J. Anim. Sci. **41,** 760–770.

MILK MARKETING BOARD, 1984: Dairy Progeny Testing Scheme (Friesians/Holstein). Report of the Breeding and Production Organisation of the Milk Marketing Board of England and Wales 34, 45.

THOMPSON, R., 1969: Iterative estimation of variance components for non-orthogonal data. Biometrics **25,** 767–773.

VAN VLECK, L. D., 1966: Correlations among records of unrelated cows in the same and different year-seasons. J. Dairy Sci. **49,** 61–64.

WILKINSON, G. N.; ECKERT, S. R.; HANCOCK, T. N.; MAYO, O., 1983: Nearest neighbour (NN) analysis of field experiments. J. Roy. Stat. Soc. B **45,** 151–211.

Authors' addresses: V. P. S. CHAUHAN, Institute of Animal Sciences, Federal Institute of Technology (ETH), CH-8092 Zurich, Switzerland; R. THOMPSON, AFRC Animal Breeding Research Organisation, West Mains Road, Edinburgh EH9 3JQ, Scotland.